

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 09:19:15 ; Search time 46 Seconds  
(without alignments)  
1880.565 Million cell updates/sec

Title: US-09-978-242-1

Perfect score: 2898

Sequence: 1 MATEHVNGTGPMDTTSA.....RGPGSGARGVRAGKRGGRS 545

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_19Jun03.\*  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2898	100.0	545	20 AAY24418	Human nucleolin-11
2	2898	100.0	545	23 ABG32846	Human nucleolin-11
3	2864	98.8	579	21 AAB56837	Human prostate can
4	2863	98.8	623	21 AAY70242	Human RNA-associat
5	2863	98.8	704	22 AAG74545	Human colon cancer
6	2860	98.7	561	23 ABB77791	Amino acid sequenc
7	2806	96.8	762	22 ABG04610	Novel human diagno
8	2355	81.3	633	21 AAY94961	Human secreted pro
9	1647.5	56.8	397	22 AAG73837	Human colon cancer

10	1342.5	46.3	673	22 ABB66506	Drosophila melanog
11	1298.5	44.8	287	20 AAY60382	Human normal bladd
12	936.5	32.3	593	21 AAB43002	Human ORFX ORF2766
13	936.5	32.3	593	22 AAM40333	Human polypeptide
14	895.5	30.9	524	22 AAM40332	Human polypeptide
15	880.5	30.4	566	22 AAM38947	Human polypeptide
16	789.	27.2	161	23 ABB77790	Amino acid sequenc
17	692	23.9	197	22 AAG75683	Human colon cancer
18	666	23.0	216	21 AAB53301	Human colon cancer
19	624.5	21.5	264	20 AAY41480	Fragment of human
20	532.5	18.4	140	21 AAG01393	Human secreted pro
21	411	14.2	151	22 ABB12296	Human APOBEC-1 sti
22	411	14.2	151	22 AAM40733	Human polypeptide
23	395.5	13.6	182	22 AAM42118	Human polypeptide
24	395.5	13.6	182	22 AAM42119	Human polypeptide
25	330	11.4	124	22 ABG47622	Human liver peptid
26	330	11.4	124	22 ABB27600	Human peptide #251
27	330	11.4	124	22 ABB32770	Peptide #276 enco
28	330	11.4	124	22 ABB18253	Protein #252 enco
29	330	11.4	124	22 AAM53577	Human brain expres
30	330	11.4	124	22 AAM65958	Human bone marrow
31	330	11.4	124	22 AAM13825	Peptide #259 enco
32	330	11.4	124	22 AAM26232	Peptide #269 enco
33	330	11.4	124	22 AAM01569	Peptide #251 enco
34	330	11.4	124	23 ABG35604	Human peptide enco
35	299.5	10.3	652	22 AAE13000	Corn poly (A) bind
36	299.5	10.3	652	22 AAE13002	Glycine max poly (
37	294.5	10.2	629	23 ABG93292	C. albicans BAX-as
38	273	9.4	55	22 AAE01648	Human secreted pro
39	269	9.3	655	22 AAE13001	Rice poly (A) bind
40	266	9.2	503	21 AAG46464	Arabidopsis thalia
41	266	9.2	515	21 AAG46463	Arabidopsis thalia
42	266	9.2	547	21 AAG46482	Arabidopsis thalia
43	266	9.2	559	21 AAG46481	Arabidopsis thalia
44	266	9.2	618	21 AAG46462	Arabidopsis thalia
45	266	9.2	662	21 AAG46480	Arabidopsis thalia

#### ALIGNMENTS

RESULT 1

AA24418

ID AAY24418 standard; Protein; 545 AA.

XX AAY24418;

AC AAY24418;

XX 21-SEP-1999 (first entry)

DT Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

PT diagnosis, prevention and treatment of disorders associated with  
 PT abnormal nucleolin-like protein expression such as cancers,  
 PT autoimmune disorders and Alzheimer's disease

XX Claim 8; Fig 1; 33pp; English.

XX The present sequence represents human nucleolin-like peptide, designated  
 CC HNLN. HNLN is the main protein component in the nucleolus of eukaryotic  
 CC cells and is an essential part of ribosome biosynthesis and also plays  
 CC an important role in importing proteins to the nucleolus. HNLN may be used  
 CC in the diagnosis, prevention and treatment of disorders associated with  
 CC abnormal expression of HNLN. For example, it may be used to treat  
 CC cancers (e.g. melanoma, breast cancer and prostate cancer), autoimmune  
 CC disorders (e.g. autoimmune haemolytic anaemia and inflammatory bowel  
 CC disease) and, in particular, Alzheimer's disease. It may also be used to  
 CC study the function of the HNLN peptides, the formation (biosynthesis) of  
 CC ribosomes and the intake of proteins into the nucleolus.

XX Sequence 545 AA;

Query Match 100.0%; Score 2898; DB 20; Length 545;  
 Best Local Similarity 100.0%; Pred. No. 7.6e-243;  
 Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MATEHVNGTGEPMDDTTSAVHSENFQTLDDAGLPQKVAEKLDEIYVAGLVAHSDLD	60
Db	1	MATEHVNGTGEPMDDTTSAVHSENFQTLDDAGLPQKVAEKLDEIYVAGLVAHSDLD	60
Qy	61	AIEALKEFNEDGALAVLQFQKSDLSHVQNKSAFLCGVMKTYRQKGTQKVDSSK	120
Db	61	AIEALKEFNEDGALAVLQFQKSDLSHVQNKSAFLCGVMKTYRQKGTQKVDSSK	120
Qy	121	EAKIKALLERTGYTLDTVTGQRKYGPPDPDSVYSGQPSVGTETIFVCKIPRDLPE	180
Db	121	EAKIKALLERTGYTLDTVTGQRKYGPPDPDSVYSGQPSVGTETIFVCKIPRDLPE	180
Qy	181	LFEXAGPIWDLRLMMDPLTGLNRYGAFVFTCTKEAAQEAQVAVKLYNNHIRSGKH	240
Db	181	LFEXAGPIWDLRLMMDPLTGLNRYGAFVFTCTKEAAQEAQVAVKLYNNHIRSGKH	240
Qy	241	ANRLFVGSIPKSKTKQILKEEFQKVTGLTDVILYHQDDKKNRGFCFLEYEDHKTAA	300
Db	241	ANRLFVGSIPKSKTKQILKEEFQKVTGLTDVILYHQDDKKNRGFCFLEYEDHKTAA	300
Qy	301	QARRLMSGKVKVNGVGTVEWADPIEDDPDVMAKVKVLFVRLNLTNTVTEILEKAFSQ	360
Db	301	QARRLMSGKVKVNGVGTVEWADPIEDDPDVMAKVKVLFVRLNLTNTVTEILEKAFSQ	360
Qy	361	FGKLERVKKLDYAFIHFDERDGAQVAMEEMNGKDLGENIEIVFAKPPDKRKAQR	420
Db	361	FGKLERVKKLDYAFIHFDERDGAQVAMEEMNGKDLGENIEIVFAKPPDKRKAQR	420
Qy	421	QAAKNQMYDDYYGPPHMPPTTRGRGGRGGYGPDPYDYDYDYDYDYDYDYDYDY	480
Db	421	QAAKNQMYDDYYGPPHMPPTTRGRGGRGGYGPDPYDYDYDYDYDYDYDYDYDY	480
Qy	481	YEDPYGYEDFQVARGRGGRGARGAAPSRCRGAAPRGRAGYSGRGGPSARGVRAGR	540
Db	481	YEDPYGYEDFQVARGRGGRGARGAAPSRCRGAAPRGRAGYSGRGGPSARGVRAGR	540
Qy	541	GRGRS 545	
Db	541	GRGRS 545	

RESULT 2  
 ABG32846

ID ABG32846 standard; Protein; 545 AA.

XX ABG32846;

XX ABG32846;

DT 18-NOV-2002 (first entry)

XX

Human nucleolin-like protein, HNLN.

Human; nucleolin-like protein; HNLN; autoimmune disorder;  
 acquired immunodeficiency syndrome; AIDS; Addison's disease; allergy;  
 asthma; atherosclerosis; multiple sclerosis; rheumatoid arthritis;  
 osteoporosis; viral infection; bacterial infection; fungal infection;  
 parasitic infection; protozoal infection; helminthic infection; cancer;  
 Alzheimer's disease; systemic sclerosis; graft-versus-host disease;  
 systemic lupus erythematosus; INCYTE 2809795.

Homo sapiens.

US2002098566-A1.

25-JUL-2002.

15-OCT-2001; 2001US-0978242.

12-DEC-1997; 97US-0990114.

01-FEB-1999; 99US-0241333.

(INCY-) INCYTE PHARM INC.

Bandman O, Yue H, Corley NC, Shah P;

WPI; 2002-690482/74.

N-PSDB; ABS52999.

Novel human nucleolin-like polypeptide, useful in diagnosis, prevention  
 and treatment of cancer, Alzheimer's disease and autoimmune disorder  
 such as AIDS, Addison's disease, allergy, asthma, and atherosclerosis -

Claim 1; Fig 1; 37pp; English.

The invention relates to an isolated human nucleolin-like polypeptide  
 (HNLN) (S1), a polypeptide comprising a naturally occurring sequence  
 having at least 90% identity to S1, or a biologically active or  
 immunogenic fragment of S1 and the HNLN encoding nucleic acid. HNLN is  
 useful for screening a compound for effectiveness as an agonist or  
 antagonist, for screening a compound that specifically binds HNLN or  
 modulates the activity of HNLN, and for preparing a polyclonal or  
 monoclonal antibody by hybridoma technology. HNLN nucleic acid is useful  
 for screening a compound for effectiveness in altering expression of a  
 target polynucleotide comprising HNLN nucleic acid and HNLN probes are  
 useful for assessing toxicity of a test compound. Anti-HNLN antibody is  
 useful in a diagnostic test for a condition or a disease associated with  
 the expression of HNLN in a biological sample, for detecting HNLN in a  
 sample, and for purifying HNLN from a sample. HNLN ant/agonists are  
 useful for treating a disease or condition associated with decreased or  
 increased expression of functional HNLN. The antibody is useful  
 for diagnosing a condition or disease associated with the expression of  
 HNLN in a subject. A HNLN nucleic acid microarray is useful for  
 generating a transcript image of a sample which contains polynucleotides.  
 HNLN and its nucleic acid are useful for diagnosing, treating and  
 preventing an autoimmune disorder (e.g. acquired immunodeficiency  
 syndrome (AIDS), Addison's disease, allergy, asthma, atherosclerosis,  
 multiple sclerosis, rheumatoid arthritis) osteoporosis, viral, bacterial,  
 fungal, parasitic, protozoal, helminthic infections, cancer,  
 Alzheimer's disease, systemic sclerosis, graft-versus-host disease and  
 systemic lupus erythematosus (many more diseases are listed in the  
 specification). The present sequence is the human HNLN protein encoded by  
 a cDNA from INCYTE clone 2809795.

Sequence 545 AA;

Query Match 100.0%; Score 2898; DB 23; Length 545;

Best Local Similarity 100.0%; Pred. No. 7.6e-243;

Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATEHVNGTGEPMDDTTSAVHSENFQTLDDAGLPQKVAEKLDEIYVAGLVAHSDLD

Db 1 MATEHVNGTGEPMDDTTSAVHSENFQTLDDAGLPQKVAEKLDEIYVAGLVAHSDLD

60

```

Best Local Similarity 100.0%; Pred. No. 6.2e-239;
Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATEHVNNGTTEPMDTTSAVIHSENFOTLLDAGLPQKVAEKLDEIYVAGLVASHDLDER 60
Db 1 MATEHVNNGTTEPMDTTSAVIHSENFOTLLDAGLPQKVAEKLDEIYVAGLVASHDLDER 60
QY 61 ATEALKEFNEDGALAVLQFQKDSLSHVQNSAFVCTKEAAQEAVALYNNHEIRSGKHIGVCISV 120
Db 61 ATEALKEFNEDGALAVLQFQKDSLSHVQNSAFVCTKEAAQEAVALYNNHEIRSGKHIGVCISV 120
QY 121 EAKIKALLERTGYTLDTVTGQKYGPPPPDSVYSGQPSVGTETPFVGIKIPRDLFEDELVP 180
Db 121 EAKIKALLERTGYTLDTVTGQKYGPPPPDSVYSGQPSVGTETPFVGIKIPRDLFEDELVP 180
QY 181 LFEKAGPIWDLRLMMDPLTGLNRGYAFVTFCTKEAAQEAVALYNNHEIRSGKHIGVCISV 240
Db 181 LFEKAGPIWDLRLMMDPLTGLNRGYAFVTFCTKEAAQEAVALYNNHEIRSGKHIGVCISV 240
QY 241 ANNRLFVGSIPKSKTKEQILEEFSKVTEGLTDVILYHQDDKKNGRGCFLYEYDHKTA 300
Db 241 ANNRLFVGSIPKSKTKEQILEEFSKVTEGLTDVILYHQDDKKNGRGCFLYEYDHKTA 300
QY 301 QARRRLMSGKVKVGNVGTVEWADPIEDDPDEVMAKVVLVFNRLANTVTTEEILEKAFSQ 360
Db 301 QARRRLMSGKVKVGNVGTVEWADPIEDDPDEVMAKVVLVFNRLANTVTTEEILEKAFSQ 360
QY 361 FGKLERVKLKDYAFIHFDERDGA VKAMEEWMNGKDEGENIEIVFAKPPDQKREKKAOR 420
Db 361 FGKLERVKLKDYAFIHFDERDGA VKAMEEWMNGKDEGENIEIVFAKPPDQKREKKAOR 420
QY 421 QAAKNQMYDDYYGYGPPHMPPTTRGRGRGGYGYPPDYGYEDYDYGYDHYNRYGG 480
Db 421 QAAKNQMYDDYYGYGPPHMPPTTRGRGRGGYGYPPDYGYEDYDYGYDHYNRYGG 480
QY 481 YEDPYGYEYDFQVARGRGGRGARGAAPSRGGAAPPRGRAGYSGRGGPGSARGVRAGKR 540
Db 481 YEDPYGYEYDFQVARGRGGRGARGAAPSRGGAAPPRGRAGYSGRGGPGSARGVRAGKR 540
QY 541 GRGRS 545
Db 541 GRGRS 545

```

RESULT 2

```

US-09-241-333-1
; Sequence 1, Application US/09241333
; Patent No. 6313266
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN NUCLEOLIN-LIKE PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/241,333
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/990,114

```

```

; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0451 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 545 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LYNNNOT05
; CLONE: 2809795
; US-09-241-333-1

```

```

Query Match 100.0%; Score 2898; DB 4; Length 545;
Best Local Similarity 100.0%; Pred. No. 6.2e-239;
Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATEHVNNGTTEPMDTTSAVIHSENFOTLLDAGLPQKVAEKLDEIYVAGLVASHDLDER 60
Db 1 MATEHVNNGTTEPMDTTSAVIHSENFOTLLDAGLPQKVAEKLDEIYVAGLVASHDLDER 60
QY 61 ATEALKEFNEDGALAVLQFQKDSLSHVQNSAFVCTKEAAQEAVALYNNHEIRSGKHIGVCISV 120
Db 61 ATEALKEFNEDGALAVLQFQKDSLSHVQNSAFVCTKEAAQEAVALYNNHEIRSGKHIGVCISV 120
QY 121 EAKIKALLERTGYTLDTVTGQKYGPPPPDSVYSGQPSVGTETPFVGIKIPRDLFEDELVP 180
Db 121 EAKIKALLERTGYTLDTVTGQKYGPPPPDSVYSGQPSVGTETPFVGIKIPRDLFEDELVP 180
QY 181 LFEKAGPIWDLRLMMDPLTGLNRGYAFVTFCTKEAAQEAVALYNNHEIRSGKHIGVCISV 240
Db 181 LFEKAGPIWDLRLMMDPLTGLNRGYAFVTFCTKEAAQEAVALYNNHEIRSGKHIGVCISV 240
QY 241 ANNRLFVGSIPKSKTKEQILEEFSKVTEGLTDVILYHQDDKKNGRGCFLYEYDHKTA 300
Db 241 ANNRLFVGSIPKSKTKEQILEEFSKVTEGLTDVILYHQDDKKNGRGCFLYEYDHKTA 300
QY 301 QARRRLMSGKVKVGNVGTVEWADPIEDDPDEVMAKVVLVFNRLANTVTTEEILEKAFSQ 360
Db 301 QARRRLMSGKVKVGNVGTVEWADPIEDDPDEVMAKVVLVFNRLANTVTTEEILEKAFSQ 360
QY 361 FGKLERVKLKDYAFIHFDERDGA VKAMEEWMNGKDEGENIEIVFAKPPDQKREKKAOR 420
Db 361 FGKLERVKLKDYAFIHFDERDGA VKAMEEWMNGKDEGENIEIVFAKPPDQKREKKAOR 420
QY 421 QAAKNQMYDDYYGYGPPHMPPTTRGRGRGGYGYPPDYGYEDYDYGYDHYNRYGG 480
Db 421 QAAKNQMYDDYYGYGPPHMPPTTRGRGRGGYGYPPDYGYEDYDYGYDHYNRYGG 480
QY 481 YEDPYGYEYDFQVARGRGGRGARGAAPSRGGAAPPRGRAGYSGRGGPGSARGVRAGKR 540
Db 481 YEDPYGYEYDFQVARGRGGRGARGAAPSRGGAAPPRGRAGYSGRGGPGSARGVRAGKR 540
QY 541 GRGRS 545
Db 541 GRGRS 545

```

RESULT 3

```

US-09-347-833-2
; Sequence 2, Application US/09347833
; Patent No. 6294658
; GENERAL INFORMATION:
; APPLICANT: Famodu, Layo O.
; APPLICANT: Odell, Joan T.
; TITLE OF INVENTION: Factors Involved in Gene Expression

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 09:23:45 ; Search time 21 Seconds  
(without alignments)  
1098.068 Million cell updates/sec

Title: US-09-978-242-1  
Perfect score: 2898  
Sequence: 1 MATEHVNGNGTEPMDTTSA.....RCGPGSARGVRACKGRGRS 545

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Issued Patents AA:\*
- 1: /cgn2\_6/ptodata/1/iaa/5A COMB.pcp.\*
  - 2: /cgn2\_6/ptodata/1/iaa/5B COMB.pcp.\*
  - 3: /cgn2\_6/ptodata/1/iaa/6A COMB.pcp.\*
  - 4: /cgn2\_6/ptodata/1/iaa/6B COMB.pcp.\*
  - 5: /cgn2\_6/ptodata/1/iaa/PCTUS COMB.pcp.\*
  - 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2898	100.0	545	2	US-08-990-114-1
2	2898	100.0	545	4	US-09-241-333-1
3	299.5	10.3	652	3	US-09-347-833-2
4	269	9.3	655	3	US-09-347-833-4
5	258.5	8.9	714	2	US-08-990-114-3
6	258.5	8.9	714	4	US-09-241-333-3
7	223.5	7.7	336	1	US-07-667-276A-8
8	209	7.2	359	1	US-07-881-075-2
9	209	7.2	359	1	US-08-120-827-2
10	209	7.2	359	1	US-08-478-675-2
11	206.5	7.1	380	1	US-07-881-075-51
12	206.5	7.1	380	1	US-08-120-827-51
13	206.5	7.1	380	1	US-08-478-675-51
14	205.5	7.1	675	3	US-08-973-273-5
15	204.5	7.1	688	3	US-08-973-273-26
16	204.5	7.1	747	3	US-08-973-273-3
17	197.5	6.8	341	2	US-08-538-711A-8
18	197.5	6.8	341	3	US-08-725-027-8
19	197.5	6.8	341	4	US-09-542-552-8
20	197.5	6.8	353	2	US-08-538-711A-7
21	197.5	6.8	353	3	US-08-725-027-7
22	197.5	6.8	353	4	US-09-542-552-7
23	197.5	6.8	444	1	US-07-881-075-3
24	197.5	6.8	444	1	US-08-120-827-3
25	197.5	6.8	444	1	US-08-478-675-3
26	194.5	6.7	223	1	US-07-667-276A-10
27	189	6.5	428	3	US-09-347-833-6

28	184	6.3	437	2	US-08-935-450-8	Sequence 8, Appli
29	181.5	6.3	323	1	US-07-667-276A-7	Sequence 7, Appli
30	181.5	6.3	485	1	US-07-881-075-1	Sequence 1, Appli
31	181.5	6.3	485	1	US-08-120-827-1	Sequence 1, Appli
32	181.5	6.3	485	1	US-08-478-675-1	Sequence 1, Appli
33	181	6.2	759	1	US-08-676-967-1	Sequence 1, Appli
34	181	6.2	759	1	US-08-676-974-1	Sequence 1, Appli
35	181	6.2	759	2	US-09-098-487-1	Sequence 1, Appli
36	179.5	6.2	284	1	US-08-390-858B-36	Sequence 36, Appli
37	177	6.1	656	2	US-08-343-443B-2	Sequence 2, Appli
38	177	6.1	656	3	US-09-214-564A-4	Sequence 4, Appli
39	169.5	5.8	471	3	US-08-866-928B-1	Sequence 1, Appli
40	169.5	5.8	471	4	US-09-685-836-1	Sequence 1, Appli
41	169	5.7	287	1	US-08-146-421-5	Sequence 5, Appli
42	166.5	5.7	414	1	US-07-667-276A-4	Sequence 4, Appli
43	158.5	5.5	217	1	US-08-390-858B-9	Sequence 9, Appli
44	147.5	5.1	162	4	US-09-575-574-4	Sequence 4, Appli
45	145.5	5.0	425	4	US-09-252-991A-17013	Sequence 17013, A

ALIGNMENTS

RESULT 1  
US-08-990-114-1  
; Sequence 1, Application US/08990114  
; Patent No. 5932475  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Yue, Henry  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Shah, Purvi  
; TITLE OF INVENTION: HUMAN NUCLEOLIN-LIKE PROTEIN  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/990,114  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0451 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 545 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: TLYNOT05  
; CLONE: 2809795  
; US-08-990-114-1  
Query Match 100.0%; Score 2898; DB 2; Length 545;

TELEFAX: 650-845-4166  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 545 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: TLYNOT05  
CLONE: 2809795  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-978-242-1

Query Match 100.0%; Score 2898; DB 9; Length 545;  
Best Local Similarity 100.0%; Pred. No. 8.3e-239; Indels 0; Gaps 0;  
Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATEHVNGTTEPMDTTSAVIHSENFQTLDDAGLPQKVAEKLDEIYVAGLVASHDLDER 60  
Db 1 MATEHVNGTTEPMDTTSAVIHSENFQTLDDAGLPQKVAEKLDEIYVAGLVASHDLDER 60

Qy 61 AIEALKEFNEDGALAVLQOFKDSLSHVQNSAFVCTKEAAQAEAVKLYNNHEIRSGKHIGVCISV 120  
Db 61 AIEALKEFNEDGALAVLQOFKDSLSHVQNSAFVCTKEAAQAEAVKLYNNHEIRSGKHIGVCISV 120

Qy 121 EAKIKALLERTGYTLDTVTGQRYGGPPDPSVYSGQPSVGTETIFVGGKIPRDLFEDELVP 180  
Db 121 EAKIKALLERTGYTLDTVTGQRYGGPPDPSVYSGQPSVGTETIFVGGKIPRDLFEDELVP 180

Qy 181 LFEKAGPIWDLRLMMDPLTGLNRGYAFVFTCTKEAAQAEAVKLYNNHEIRSGKHIGVCISV 240  
Db 181 LFEKAGPIWDLRLMMDPLTGLNRGYAFVFTCTKEAAQAEAVKLYNNHEIRSGKHIGVCISV 240

Qy 241 ANNRLFVGSIPKSTKEQILEEFSKVTEGLTDVILYHQDDKKKNGFCFLEYEDHKTAA 300  
Db 241 ANNRLFVGSIPKSTKEQILEEFSKVTEGLTDVILYHQDDKKKNGFCFLEYEDHKTAA 300

Qy 301 QARRRLMSGKVKVNGVGTVEWADPIEDDPPEVMAKVKVLFVRNLANTVTTEEILEKAFSQ 360  
Db 301 QARRRLMSGKVKVNGVGTVEWADPIEDDPPEVMAKVKVLFVRNLANTVTTEEILEKAFSQ 360

Qy 361 FGKLERVKKLDYAFIHFDERDGAVKAMEENMGKLEGENIEIVFAKPPDQKRRKAQR 420  
Db 361 FGKLERVKKLDYAFIHFDERDGAVKAMEENMGKLEGENIEIVFAKPPDQKRRKAQR 420

Qy 421 QAAKNQMYDDYYGYGPPHMPPTTRGRGRGGGYGPPDYGYEDYDYGYDHYNRYGG 480  
Db 421 QAAKNQMYDDYYGYGPPHMPPTTRGRGRGGGYGPPDYGYEDYDYGYDHYNRYGG 480

Qy 481 YEDPYGYEDFQVARGRGARGAAPSRRGGAAPPRGRAGYSORGGPGSARGVRAGKR 540  
Db 481 YEDPYGYEDFQVARGRGARGAAPSRRGGAAPPRGRAGYSORGGPGSARGVRAGKR 540

Qy 541 GRGRS 545  
Db 541 GRGRS 545

RESULT 2  
US-09-925-300-1415  
Sequence 1415, Application US/0925300  
Patent No. US20020151681A1  
GENERAL INFORMATION:  
APPLICANT: Craig Rosen,  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA101  
CURRENT APPLICATION NUMBER: US/09/925,300  
PRIOR FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05988  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270

PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1890  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1415  
LENGTH: 579  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-925-300-1415

Query Match 98.8%; Score 2864; DB 10; Length 579;  
Best Local Similarity 98.2%; Pred. No. 7.3e-236; Indels 6; Gaps 1;  
Matches 540; Conservative 2; Mismatches 2; Indels 6; Gaps 1;

Qy 1 MATEHVNGTTEPMDTTSAVIHSENFQTLDDAGLPQKVAEKLDEIYVAGLVASHDLDER 60  
Db 18 MATEHVNGTTEPMDTTSAVIHSENFQTLDDAGLPQKVAEKLDEIYVAGLVASHDLDER 77

Qy 61 AIEALKEFNEDGALAVLQOFKDSLSHVQNSAFVCTKEAAQAEAVKLYNNHEIRSGKHIGVCISV 120  
Db 78 AIEALKEFNEDGALAVLQOFKDSLSHVQNSAFVCTKEAAQAEAVKLYNNHEIRSGKHIGVCISV 137

Qy 121 EAKIKALLERTGYTLDTVTGQRYGGPPDPSVYSGQPSVGTETIFVGGKIPRDLFEDELVP 180  
Db 138 EAKIKALLERTGYTLDTVTGQRYGGPPDPSVYSGQPSVGTETIFVGGKIPRDLFEDELVP 197

Qy 181 LFEKAGPIWDLRLMMDPLTGLNRGYAFVFTCTKEAAQAEAVKLYNNHEIRSGKHIGVCISV 240  
Db 198 LFEKAGPIWDLRLMMDPLTGLNRGYAFVFTCTKEAAQAEAVKLYNNHEIRSGKHIGVCISV 257

Qy 241 ANNRLFVGSIPKSTKEQILEEFSKVTEGLTDVILYHQDDKKKNGFCFLEYEDHKTAA 300  
Db 258 ANNRLFVGSIPKSTKEQILEEFSKVTEGLTDVILYHQDDKKKNGFCFLEYEDHKTAA 317

Qy 301 QARRRLMSGKVKVNGVGTVEWADPIEDDPPEVMAKVKVLFVRNLANTVTTEEILEKAFSQ 360  
Db 318 QARRRLMSGKVKVNGVGTVEWADPIEDDPPEVMAKVKVLFVRNLANTVTTEEILEKAFSQ 377

Qy 361 FGKLERVKKLDYAFIHFDERDGAVKAMEENMGKLEGENIEIVFAKPPDQKRRKAQR 420  
Db 378 FGKLERVKKLDYAFIHFDERDGAVKAMEENMGKLEGENIEIVFAKPPDQKRRKAQR 437

Qy 421 QAAKNQMYDDYYGYGPPHMPPTTRGRGRGGGYGPPDYGYEDYDYGYDHYNRYGG 480  
Db 438 QAAKNQMYDDYYGYGPPHMPPTTRGRGRGGGYGPPDYGYEDYDYGYDHYNRYGG 497

Qy 481 YEDPYGYEDFQVARGRGARGAAPSRRGGAAPPRGRAGYSORGGPGSARGVRAGK - 539  
Db 498 YEDPYGYEDFQVARGRGARGAAPSRRGGAAPPRGRAGYSORGGPGSARGVRAGK 557

Qy 540 -----RGRGR 544  
Db 558 GAQOQRGRGQ 567

RESULT 3  
US-10-106-698-5319  
Sequence 5319, Application US/10106698  
Publication No. US20030109690A1  
GENERAL INFORMATION:  
APPLICANT: Ruben et al.  
TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides  
FILE REFERENCE: PA005P1  
CURRENT APPLICATION NUMBER: US/10/106,698  
CURRENT FILING DATE: 2002-03-27  
PRIOR APPLICATION NUMBER: PCT/US00/26524  
PRIOR FILING DATE: 2000-09-28  
PRIOR APPLICATION NUMBER: US 60/157,137  
PRIOR FILING DATE: 1999-09-29  
PRIOR APPLICATION NUMBER: US 60/163,280  
PRIOR FILING DATE: 1999-11-03  
NUMBER OF SEQ ID NOS: 8564  
SOFTWARE: PatentIn Ver. 3.0  
SEQ ID NO 5319

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 09:26:11 ; Search time 38 Seconds  
(without alignments)  
2965.039 Million cell updates/sec

Title: US-09-978-242-1

Perfect score: 2898

Sequence: 1 MATEHVNGNGTEPMDTTSA.....RGPGSARGVRAGKRGGRS 545

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 777136 seqs, 206736638 residues

Total number of hits satisfying chosen parameters: 777136

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	2898	100.0	545	9	US-09-978-242-1
2	2864	98.8	579	10	US-09-925-300-1415
3	2863	98.8	704	15	US-10-106-698-5319
4	2860	98.7	561	10	US-09-821-687-4
5	2355	81.3	633	10	US-09-821-687-10
6	2355	81.3	633	11	US-09-374-046A-128
7	1647.5	56.8	397	15	US-10-106-698-4611
8	1342.5	46.3	673	12	US-09-949-029-100
9	1075	37.1	335	10	US-09-821-687-11
10	789	27.2	161	10	US-09-821-687-2
11	692	23.9	197	15	US-10-106-698-6457
12	666	23.0	216	9	US-09-925-299-841
13	666	23.0	216	11	US-09-925-299-841
14	330	11.4	124	9	US-09-864-761-33551
15	266	9.2	662	12	US-10-338-777-42

258.5	8.9	714	9	US-09-978-242-3
249.5	8.6	402	14	US-10-062-254-258
249.5	8.6	687	12	US-10-104-047-2651
244	8.4	707	12	US-10-384-569-3
236.5	8.2	406	14	US-10-062-254-238
234.5	8.1	409	14	US-10-062-254-276
233	8.0	314	9	US-09-849-967A-3
233	8.0	320	12	US-10-341-434-59
232	8.0	428	14	US-10-062-254-250
228.5	7.9	430	14	US-10-062-254-252
227	7.8	332	12	US-10-108-260A-4506
226.5	7.8	633	12	US-10-341-434-49
225	7.8	301	9	US-09-799-777-64
221.5	7.6	432	14	US-10-062-254-242
217	7.5	397	14	US-10-062-254-268
216	7.5	378	9	US-09-849-967A-2
214.5	7.4	436	14	US-10-062-254-248
209	7.2	359	14	US-10-025-367-23
208	7.2	416	14	US-10-062-254-264
207.5	7.2	350	14	US-10-062-254-260
206.5	7.1	380	14	US-10-025-367-27
206.5	7.1	420	14	US-10-062-254-254
205	7.1	117	9	US-09-764-887-194
205	7.1	117	15	US-10-073-961-194
202	7.0	494	9	US-09-833-790-234
201	6.9	356	14	US-10-062-254-262
197.5	6.8	353	11	US-09-935-642-14
194.5	6.7	359	14	US-10-025-367-25
193.5	6.7	316	14	US-10-062-254-266
192.5	6.6	307	14	US-10-062-254-244

#### ALIGNMENTS

#### RESULT 1

US-09-978-242-1  
; Sequence 1, Application US/09978242  
; Patent No. US20020098566A1  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; Yue, Henry  
; Corley, Neil C.  
; Shah, Purvi  
; TITLE OF INVENTION: HUMAN NUCLEOLIN-LIKE PROTEIN  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION NUMBER: US/09/978,242  
; FILING DATE: 15-Oct-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/241,333  
; FILING DATE: <Unknown>  
; APPLICATION NUMBER: 08/990,114  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0451 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 09:23:10 ; Search time 20 Seconds  
(without alignments)  
2620.597 Million cell updates/sec

Title: US-09-978-242-1  
Perfect score: 2898  
Sequence: 1 MATEHVNGNGTEPMDTTSA.....RGGPSRGVRGKRGGRS 545  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues  
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_76:.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2355	81.3	633	2 T02673	heterogeneous nucl
2	495.5	17.1	521	2 T01563	hypothetical prote
3	435.5	15.0	471	2 T49019	probable RNA bindi
4	287.5	9.9	653	1 DNZPPA	polyadenylate-bind
5	286	9.9	651	2 S18874	nucleolin - Africa
6	280	9.7	693	2 JC7925	nucleolin - common
7	279	9.6	671	2 C96534	probable Poly-A Bi
8	278.5	9.6	705	2 S32644	nucleolin - Africa
9	275	9.5	345	1 B41732	heterogeneous nucl
10	274	9.5	651	2 T06979	polyadenylate-bind
11	268.5	9.3	623	2 T07933	polyadenylate-bind
12	266	9.2	662	2 T00497	polyadenylate-bind
13	258.5	8.9	629	2 T05425	polyadenylate-bind
14	258.5	8.9	713	2 A27441	nucleolin - Chinae
15	253	8.7	405	2 H86249	hypothetical prote
16	251.5	8.7	638	2 S37085	polyadenylate-bind
17	251	8.7	448	2 T15542	hypothetical prote
18	249	8.6	353	1 S56750	single stranded D
19	249	8.6	692	2 T21095	hypothetical prote
20	247	8.5	500	2 S55785	nucleolar protein
21	244	8.4	707	2 A35804	nucleolin - human
22	242.5	8.4	636	2 T48718	poly(A) binding pr
23	242	8.4	427	2 T04823	hypothetical prote
24	241.5	8.3	712	2 JH0148	nucleolin - rat
25	239	8.2	320	2 S02061	heterogeneous ribo
26	236.5	8.2	668	2 B96740	hypothetical prote
27	236	8.1	577	1 DNBPFA	polyadenylate-bind
28	234.5	8.1	574	2 S30887	polyadenylate-bind
29	233	8.0	320	1 DDRT	helix-destabilizin

RESULT 1

T02673

heterogeneous nuclear ribonucleoprotein R - human

C;Species: Homo sapiens (man)

C;Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 08-Oct-1999

C;Accession: T02673

R;Chan, E.K.L.; Mathison, D.A.; Portman, D.; Dreyfuss, G.; Steiner, G.; Tan, E.M.; Hassfe

Nucleic Acids Res. 26, 439-445, 1998

A;Title: Molecular definition of heterogeneous nuclear ribonucleoprotein R (hnRNP R) usir

A;Reference number: Z14697; MUID:98083170; PMID:9421497

A;Accession: T02673

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-633 <CHA>

A;Cross-references: EMBL:AF000364; NID:g2697102; PIDN:AAC39540.1; PID:g2697103

Query Match 81.3%; Score 2355; DB 2; Length 633;  
Best Local Similarity 80.2%; Pred. No 5.1e-143;  
Matches 449; Conservative 51; Mismatches 40; Indels 20; Gaps 10;

QY 1 MATEHVNGNGT-----EEPMDTTSAVIHSENFQTLTDLGLPQKVAEKDLIYVAGLVAHS 55  
DB 1 MANQ-VNGNAVOLKEEPEMD-TSSVTTHYKTLIEAGLPQKVAERLDEIFQTLGVAHV 58  
QY 56 DLDERAIEALKKEPNEGCALAVLQOFKDSLSHVONKSAPFLCGVMKTYRQREKQTKVADS 115  
DB 59 DLDERAIDALREFNEEGALSVLQOFKESDLSHVONKSAPFLCGVMKTYRQREKQSKVQES 118  
QY 116 SKGPDSEAKTKALLERTGYTLDTVTGORKYGGPPPSVYSGQPSVGTETIFVGKIPRDLFE 175  
DB 119 TKGPDSEAKTKALLERTGYTLDTVTGORKYGGPPPSVYSGQPSVGTETIFVGKIPRDLFE 178  
QY 176 DELVPLFEKAGPIWDLRLMMDPLTGLNRYGAFVFTCTKEAAQEAQVAVKLYNNHIRSKHG 235  
DB 179 DELVPLFEKAGPIWDLRLMMDPLSGQNRGYAFITFCGEAAQEAQVAVKLYNNHIRSKHG 238  
QY 236 VCISVANNRLFVGSIPKSKTEQILKEPSKVTEGLTDVILYHQDDKKNNRQGFCELEYED 295  
DB 239 VCISVANNRLFVGSIPKSKTEQILKEPSKVTEGLTDVILYHQDDKKNNRQGFCELEYED 298  
QY 296 HKTAAQARRLMSGKVKVNGVGTVEWADPIEDPPEVMAKVKVLFVRNLANTVTTEEILE 355  
DB 299 HKSAAQARRLMSGKVKVNGVGTVEWADPIEDPPEVMAKVKVLFVRNLANTVTTEEILE 358  
QY 356 KAFSGFKGLRVKLDYAFTHFDERDGAVKAMEENMGKDEGENIEIIVFAKPPDQKRKE 415  
DB 359 KSFSFEGKGLRVKLDYAFVHFDEGAAGAAVKAMDEMGKEIEGEEIEIIVLAKPPDQKRKE 418  
QY 416 RQAQQAQKQYDDYYGGPPHPPPTGRGR-GRGGYGYPDYGYVEDYY-DYGYVD 473  
DB 419 ROAARQASRSTAYEDYYHHPPPMPPPTGRGRGGRGGYGYPPDYGYEDYYDDYYGYD 478



```

FT DOMAIN 412 418 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL) .
FT DOMAIN 165 244 RNA-BINDING (RRM) 1.
FT DOMAIN 246 328 RNA-BINDING (RRM) 2.
FT DOMAIN 341 411 RNA-BINDING (RRM) 3.
FT DOMAIN 447 567 RNA-BINDING (RGG-BOX) .
FT DOMAIN 462 471 3 X 11 AA APPROXIMATE REPEATS OF D-D-Y-Y-
FT REPEAT 472 482 G-Y-D-Y-H-D-Y.
FT REPEAT 488 497 1 (APPROXIMATE) .
FT DOMAIN 579 633 GLN/ASN-RICH DOMAIN.
FT SEQUENCE 633 AA; 70943 MW; 088341F6465ED46F CRC64;

Query Match
Best Local Similarity 80.28; Score 2355; DB 1; Length 633;
Matches 449; Conservative 51; Mismatches 40; Indels 20; Gaps 10;

Qy 1 MATHVNGNT-----EPMDDTTSVTHSENFTLLDAGLPQKVAELDIYVAGLVAHS 55
Db 1 MANQ-VNGNAVQLKEEPMDD-TSSVTHTEHYKTLIEAGLPQKVAELDIYVAGLVAHV 58
Qy 56 DLDERATEALKEFNEDGALVLOQFSDLSHVQNSAFVLCGVNKTYYRQREKQTKVADS 115
Db 59 DLDERATDALREFNEGALSLSVLOQFSDLSHVQNSAFVLCGVNKTYYRQREKQSKVOES 118
Qy 116 SKGPDEAKIKALLERTGYTLDTVTGQKYGPPDPDSVYSGOQPSVGTETIFVGKIPDLFE 175
Db 119 TKGPDEAKIKALLERTGYTLDTVTGQKYGPPDPDSVYSGVQPGIGTEVFVGKIPDLFE 178
Qy 176 DELVPLFEKAGPIWDLRLMDPLTGLNRGAFVFTCTKEAAQEAVALYNNHIEIRSGKHG 235
Db 179 DELVPLFEKAGPIWDLRLMDPLTGLNRGAFVFTCTKEAAQEAVALYNNHIEIRSGKHG 238
Qy 236 VCISVANNRLPVGSIPIKSTKEQILEEFSKVTGELTDVILYHQDDKKQKRGFCFLBYED 295
Db 239 VCISVANNRLPVGSIPIKSTKEQILEEFSKVTGELTDVILYHQDDKKQKRGFCFLBYED 298
Qy 296 HKTAQAARRRLMSKVKVGNVGVTEWADPTEDPPEVMAKVLFVFNLIANTVTEILE 355
Db 299 HKSAQAARRRLMSKVKVGNVGVTEWADPVEEDPPEVMAKVLFVFNLIANTVTEILE 358
Qy 356 KAFSQFQKLERVKKLDYAFIHFDERDGAVKAMEMNGKOLEGENIEIVFAKPPDKRKE 415
Db 359 KSFSEFQKLERVKKLDYAFVHFDERDGAVKAMEMNGKOLEGENIEIVFAKPPDKRKE 418
Qy 416 RKAORQAANKQMDYDYIYGGPHPPPTPRGRGR-GRGGCGYGPDPDYGYEDYY-DYGYD 473
Db 419 RQAARQASRSTAYEDYYHPPPPPTPRGRGRGRGGGCGYGPDPDYGYEDYYDYDYD 478
Qy 474 YHNYRGVYEDPYGYED-FQVGARGRGGRGARGA-APSRGRGAAPPGRAGYSQGGP-G 530
Db 479 YHNYRGVYEDPYGYEDDGYAVRGH-GRGGRGAPPPPRGRGAPPPGRAGYSQGGP 537
Qy 531 SARGVRAGK-----RRRG 543
Db 538 PPRGSRGGRGGGPAQOQGRG 557

RESULT 2
PABP_SCHPO STANDARD; PRT; 653 AA.
AC P31209; P87135;
DT 01-JUL-1993 (Rel. 26, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Polyadenylate-binding protein (Poly(A)-binding protein) (PABP) .
GN PAB1 OR PABP OR SPAC57A7.04C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetaceae.
OX NCBI_TaxID=4896;
RN [1]

```

```

SEQUENCE FROM N.A.
RP STRAIN=972;
RC MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skellon J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Welfens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritsch C., Holzer B., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Furnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Dominguez A., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Daga R.R., Revuelta J.L., Moreno S., Armstrong J., Foreburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RN Nature 415:871-880(2002) .
RP [2]
RC SEQUENCE OF 15-628 FROM N.A.
RA MEDLINE=91260690; PubMed=1675426;
RA Burd C.G., Matunis E.L., Dreyfuss G.;
RT "The multiple RNA-binding domains of the mRNA poly(A)-binding protein
RT have different RNA-binding activities.";
RL Mol. Cell. Biol. 11:3419-3424(1991) .
CC -|- FUNCTION: BINDS THE POLY(A) TAIL OF MRNA.
CC -|- SUBCELLULAR LOCATION: Cytoplasmic.
CC -|- SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z95396; CAB08762.1; -
DR EMBL; M64603; AAA35320.1; -
DR PIR; T38950; DNZPPA.
DR HSP; P11940; 1CVJ
DR GeneDB_Spombe; SPAC57A7.04c; -
DR InterPro; IPR002004; PABP/HECT.
DR InterPro; IPR006515; PABP_1234.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00658; PABP; 1.
DR Pfam; PF00076; Rtm; 4.
DR SMART; SM00517; PolyA; 1.
DR SMART; SM00360; RRM; 4.
DR TIGRFAMs; TIGR01628; PABP-1234; 1.
DR PROSITE; PS0102; RRM; 4.
DR PROSITE; PS00030; RRM_RNP_1; 3.
KW RNA-binding; Repeat.
FT DOMAIN 80 158 RNA-BINDING (RRM) 1.
FT DOMAIN 168 245 RNA-BINDING (RRM) 2.
FT DOMAIN 261 338 RNA-BINDING (RRM) 3.
FT DOMAIN 364 441 RNA-BINDING (RRM) 4.
FT CONFLICT 15 34 ESDVNTNNEAVESSTKEES -> MSLENSSTLSCSNNT
FT CONFLICT 349 349 R -> A (IN REF. 2).
FT CONFLICT 518 526 TQFPAGGPA -> PLSSLLVLR (IN REF. 2).
FT SEQUENCE 653 AA; 71512 MW; 7F8F5CAD69D0CFE1 CRC64;

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 09:19:30 ; Search time 17 Seconds  
(without alignments)  
1507.621 Million cell updates/sec

Title: US-09-978-242-1

Perfect score: 2898

Sequence: 1 MATEHVNGNGTEPMDTTSA.....RGGPSARGVYRACKRGGRS 545

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2355	81.3	633	1 ROR HUMAN	O43390 homo sapien
2	287.5	9.9	653	1 PABP SCHPO	P13290 schizosacch
3	286	9.9	650	1 NUCLE XENLA	P20397 xenopus lae
4	274	9.5	345	1 SOD DROME	O08473 drosophila
5	260.5	9.0	644	1 PAB4 HUMAN	O13310 homo sapien
6	258.5	8.9	629	1 PAB2 ARATH	P42731 arabidopsis
7	258.5	8.9	713	1 NUCLE MESAU	P08199 mesocricetu
8	248	8.6	631	1 PAB3 HUMAN	Q9h361 homo sapien
9	247	8.5	500	1 GAR2 SCHPO	P19338 homo sapien
10	244	8.4	706	1 NUCLE HUMAN	P19338 homo sapien
11	242.5	8.4	636	1 PAB1 MOUSE	P29341 mus musculu
12	241.5	8.3	712	1 NUCLE RAT	P13383 rattus norv
13	240.5	8.3	636	1 PAB1 ARATH	P11940 homo sapien
14	236.5	8.2	668	1 PAB5 ARATH	P15196 arabidopsis
15	236	8.1	576	1 PABP YEAST	P04147 saccharomyc
16	234.5	8.1	632	1 PABP DROME	P21187 drosophila
17	233	8.0	319	1 ROAL MOUSE	P49312 mus musculu
18	233	8.0	706	1 NUCLE MOUSE	P09405 mus musculu
19	232.5	8.0	414	1 NOP3 YEAST	Q01560 saccharomyc
20	232.5	8.0	633	1 PAB1 XENLA	P20965 xenopus lae
21	231	8.0	319	1 ROAL RAT	P04256 rattus norv
22	229	7.9	391	1 ROG HUMAN	P38159 homo sapien
23	229	7.9	694	1 NUCLE CHICK	P15771 gallus gall
24	227	7.8	522	1 PAB2 HUMAN	P15097 homo sapien
25	224.5	7.7	660	1 PAB3 ARATH	O64380 arabidopsis
26	223	7.7	319	1 ROAL MACMU	Q28521 macaca mula
27	222.5	7.7	353	1 ROD RAT	O91154 rattus norv
28	222	7.7	371	1 ROAL HUMAN	P09651 homo sapien
29	221.5	7.6	379	1 ROAL MOUSE	Q8b905 mus musculu
30	221	7.6	609	1 PABX ARATH	O92q48 arabidopsis
31	219.5	7.6	305	1 ROAO HUMAN	Q13151 homo sapien
32	219.5	7.6	355	1 ROD HUMAN	Q14103 homo sapien
33	218.5	7.5	424	1 S3B4_HUMAN	Q15427 homo sapien

#### ALIGNMENTS

##### RESULT 1

ID	ROR_HUMAN	STANDARD;	PRT;	633 AA.
AC	O43390;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Heterogeneous nuclear ribonucleoprotein R (hnRNP R).			
GN	HNPRP.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=980831170; PubMed=9421497;			
RA	Hassfeld W., Chan E.K.L., Mathison D.A., Portman D., Dreyfuss G.,			
RA	Steiner G., Tan E.M.;			
RT	"Molecular definition of heterogeneous nuclear ribonucleoprotein R			
RT	(hnRNP R) using autoimmune antibody: Immunological relationship with			
RT	hnRNP P.";			
RL	Nucleic Acids Res. 26:439-445(1998).			
CC	-1- FUNCTION: COMPONENT OF RIBONUCLEOSOMES, WHICH ARE COMPLEXES OF AT			
CC	LEAST 20 OTHER DIFFERENT HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEINS			
CC	(HNPRNP). HNPRNP PLAY AN IMPORTANT ROLE IN PROCESSING OF PRECURSOR			
CC	MRNA IN THE NUCLEUS.			
CC	-1- SUBCELLULAR LOCATION: Nuclear; nucleoplasm.			
CC	-1- SIMILARITY: Contains 3 RNA recognition motif (RNM) domains.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; AF000364; AAC39540.1; --			
DR	PIR; T02673; T02673.			
DR	HSP; P09651; IHA1.			
DR	Genew; HGNC:5047; HNPRP.			
DR	GK; O43390; --			
DR	MIM; 607201; --			
DR	GO; GO:0005634; C:nucleus; TAS.			
DR	GO; GO:0008436; F:heterogeneous nuclear ribonucleoprotein; TAS.			
DR	GO; GO:0003723; F:RNA binding activity; TAS.			
DR	GO; GO:0006397; P:mrna processing; TAS.			
DR	InterPro; IPR000504; hnRNP_R_Q.			
DR	InterPro; IPR000535; hnRNP_R_Q.			
DR	Pfam; PF00076; rrm; 3.			
DR	SMART; SM00360; RRM; 3.			
DR	TIGRFAM; TIGR01648; hnRNP-R-Q; 1.			
DR	PROSITE; PS0102; RRM; 3.			
DR	PROSITE; PS00030; RRM RNP 1; 2.			
KW	Nuclear protein; RNA-binding; Ribonucleoprotein; Repeat.			
FT	DOMAIN 1 153 ASP/GLU-RICH (ACIDIC).			

P51992 xenopus lae  
O97018 chrysomya r  
O35479 mus musculu  
P51968 xenopus lae  
O17310 musca domes  
Q01085 homo sapien  
P51989 xenopus lae  
Q96du9 homo sapien  
Q12926 homo sapien  
P51990 xenopus lae  
Q60899 mus musculu  
O09032 rattus norv

34 218 7.5 385 1 RO32 XENLA  
35 215.5 7.4 307 1 SXL CHRRU  
36 214.5 7.4 388 1 ROG MOUSE  
37 214 7.4 373 1 RO31 XENLA  
38 212.5 7.3 324 1 SXL MUSDO  
39 212 7.3 375 1 TIAR HUMAN  
40 210 7.2 346 1 RO21 XENLA  
41 209.5 7.2 382 1 PAB5 HUMAN  
42 209 7.2 359 1 ELV2 HUMAN  
43 208 7.2 358 1 RO22 XENLA  
44 207.5 7.2 360 1 ELV2 MOUSE  
45 206.5 7.1 373 1 ELV4 RAT

```

Db 121 EAKIKALLERTGYTLDTVTGQKYGPPDSVYSGQPPSVGTEIFVGKIPRDLFEDELVP 180
Qy 181 LFEKAGPIWDLRLMDPLTGLNRGYAFVFTCTKEAAQEAVKLYNNHIEIRSGKHIGVCISV 240
Db 181 LFEKAGPIWDLRLMDPLTGLNRGYAFVFTCTKEAAQEAVKLYNNHIEIRSGKHIGVCISV 240
Qy 241 ANNRLFVGSIPKSKTKEQILEEFSKVTGLTDVILYHQDDKKKNGFCFLEVEDHKTAA 300
Db 241 ANNRLFVGSIPKSKTKEQILEEFSKVTGLTDVILYHQDDKKKNGFCFLEVEDHKTAA 300
Qy 301 QARRRLMSGKVKVGNVGTVEWADPIEDDPEVMAKVVLFRNLANTVTTEILEKAFSQ 360
Db 301 QARRRLMSGKVKVGNVGTVEWADPIEDDPEVMAKVVLFRNLANTVTTEILEKAFSQ 360
Qy 361 FGKLERVKKLDYAFIHFDERDGAVKAMEEMNGKDLGENIEIVFAKPPDQKPKERKAQR 420
Db 361 FGKLERVKKLDYAFIHFDERDGAVKAMEEMNGKDLGENIEIVFAKPPDQKPKERKAQR 420
Qy 421 QAAKNQMYDDYVYGGPPHMPPTTRGRGRGGGYPDPDYGYEDYDYDYGYDYNHNRGG 480
Db 421 QAAKNQMYDDYVYGGPPHMPPTTRGRGRGGGYPDPDYGYEDYDYDYGYDYNHNRGG 480
Qy 481 YEDPYGYEDFQVARGRGARGAAPSRRGGAAPPRGRAGYSQGGPGSARGVRGARG 539
Db 481 YEDPYGYEDFQVARGRGARGAAPSRRGGAAPPRGRAGYSQGGPGSARGVRGARG 540
Qy 540 -----RGRG 544
Db 541 GAOQQRGRG 550

```

RESULT 2

```

Q9QYF4 ID Q9QYF4 PRELIMINARY; PRT; 561 AA.
AC Q9QYF4
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE SYNCRIP protein.
GN NSAP1 OR SYNCRIP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DDY;
RX MEDLINE=20200483; PubMed=10734137;
RA Mizutani A., Fukuda M., Iwata K., Shiraiishi Y., Mikoshiba K.;
RT "SYNCRIP, a cytoplasmic counterpart of heterogeneous nuclear
RT ribonucleoprotein R, interacts with ubiquitous synaptotagmin
RT isoforms.";
RL J. Biol. Chem. 275:9823-9831(2000).
DR EMBL; AB035725; BAA88342.1; -.
DR HSP; P19339; 2SXL.
DR MGD; MGI:1891690; Neap1.
DR InterPro; IPR006535; hRNP_R_Q.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 3.
DR SMART; SM00360; RRM; 3.
DR TIGRFAMs; TIGR01648; hRNP-R-Q; 1.
DR PROSITE; PS0102; RRM; 3.
DR PROSITE; PS00030; RRM_RNP_1; 2.
SQ SEQUENCE 561 AA; 62544 MW; C0259C340146D16A CRC64;

```

```

Query Match 98.7%; Score 2860; DB 11; Length 561;
Best Local Similarity 98.2%; Pred. No. 1.7e-191;
Matches 539; Conservative 2; Mismatches 2; Indels 6; Gaps 1;

Qy 1 MATEVNGNGTEPMDTTSVAVHSENFQTLDDAGLPQKVAEKLDEIYVAGLVVHSDILDER 60
Db 1 MATEVNGNGTEPMDTTSVAVHSENFQTLDDAGLPQKVAEKLDEIYVAGLVVHSDILDER 60

```

```

Qy 61 AIEALKEFNEDGALAVLQOFKDSLSHVQNKSAFLCGVMKTYRQREKQGTKVADSSKGP 120
Db 61 AIEALKEFNEDGALAVLQOFKDSLSHVQNKSAFLCGVMKTYRQREKQGTKVADSSKGP 120
Qy 121 EAKIKALLERTGYTLDTVTGQKYGPPDSVYSGQPPSVGTEIFVGKIPRDLFEDELVP 180
Db 121 EAKIKALLERTGYTLDTVTGQKYGPPDSVYSGQPPSVGTEIFVGKIPRDLFEDELVP 180
Qy 181 LFEKAGPIWDLRLMDPLTGLNRGYAFVFTCTKEAAQEAVKLYNNHIEIRSGKHIGVCISV 240
Db 181 LFEKAGPIWDLRLMDPLTGLNRGYAFVFTCTKEAAQEAVKLYNNHIEIRSGKHIGVCISV 240
Qy 241 ANNRLFVGSIPKSKTKEQILEEFSKVTGLTDVILYHQDDKKKNGFCFLEVEDHKTAA 300
Db 241 ANNRLFVGSIPKSKTKEQILEEFSKVTGLTDVILYHQDDKKKNGFCFLEVEDHKTAA 300
Qy 301 QARRRLMSGKVKVGNVGTVEWADPIEDDPEVMAKVVLFRNLANTVTTEILEKAFSQ 360
Db 301 QARRRLMSGKVKVGNVGTVEWADPIEDDPEVMAKVVLFRNLANTVTTEILEKAFSQ 360
Qy 361 FGKLERVKKLDYAFIHFDERDGAVKAMEEMNGKDLGENIEIVFAKPPDQKPKERKAQR 420
Db 361 FGKLERVKKLDYAFIHFDERDGAVKAMEEMNGKDLGENIEIVFAKPPDQKPKERKAQR 420
Qy 421 QAAKNQMYDDYVYGGPPHMPPTTRGRGRGGGYPDPDYGYEDYDYDYGYDYNHNRGG 480
Db 421 QAAKNQMYDDYVYGGPPHMPPTTRGRGRGGGYPDPDYGYEDYDYDYGYDYNHNRGG 480
Qy 481 YEDPYGYEDFQVARGRGARGAAPSRRGGAAPPRGRAGYSQGGPGSARGVRGARG 539
Db 481 YEDPYGYEDFQVARGRGARGAAPSRRGGAAPPRGRAGYSQGGPGSARGVRGARG 540
Qy 540 -----RGRG 543
Db 541 GAOQQRGRG 549

```

RESULT 3

```

Q9Y583 ID Q9Y583 PRELIMINARY; PRT; 562 AA.
AC Q9Y583;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NSAP1 protein.
GN NSAP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99102562; PubMed=9847309;
RA Harris C.E., Boden R.A., Astell C.R.;
RT "A novel heterogeneous nuclear ribonucleoprotein-like protein
RT interacts with NS1 of the minute virus of mice.";
RL J. Virol. 73:72-80(1999).
DR EMBL; AF155568; AAD38198.1; -.
DR HSP; P19339; 2SXL.
DR InterPro; IPR006535; hRNP_R_Q.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 3.
DR SMART; SM00360; RRM; 3.
DR TIGRFAMs; TIGR01648; hRNP-R-Q; 1.
DR PROSITE; PS0102; RRM; 3.
DR PROSITE; PS00030; RRM_RNP_1; 2.
SQ SEQUENCE 562 AA; 62656 MW; D48BE582B00F946D CRC64;

```

```

Query Match 98.7%; Score 2860; DB 4; Length 562;
Best Local Similarity 98.0%; Pred. No. 1.7e-191;
Matches 539; Conservative 3; Mismatches 2; Indels 6; Gaps 1;

Qy 1 MATEVNGNGTEPMDTTSVAVHSENFQTLDDAGLPQKVAEKLDEIYVAGLVVHSDILDER 60

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 09:22:40 ; Search time 39 Seconds  
(without alignments)  
3606.120 Million cell updates/sec

Title: US-09-978-242-1

Perfect score: 2898

Sequence: 1 MATEHVNGTPEPMDTSA.....RGPGSARGVRACKRGGRS 545

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL\_23:\*
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phage:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_rvirus:\*
  - 16: sp\_bacteriap:\*
  - 17: sp\_cheap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2861	98.7	562	11 Q8BGP1	Q8bgl1 mus musculus
2	2860	98.7	561	11 Q9QYF4	Q9qyf4 mus musculus
3	2860	98.7	562	4 Q9Y583	Q9y583 homo sapien
4	2860	98.7	627	11 Q8QGC2	Q8qgc2 mus musculus
5	2857	98.6	561	4 Q96LC1	Q96lc1 homo sapien
6	2857	98.6	623	4 Q60506	Q60506 homo sapien
7	2829	97.6	625	11 Q88991	Q88991 mus musculus
8	2647.5	91.4	527	4 Q8N599	Q8n599 homo sapien
9	2640.5	91.1	588	4 Q96LC2	Q96lc2 homo sapien
10	2533	87.4	558	11 Q91ZR0	Q91zr0 mus musculus
11	2487	85.8	491	11 Q8C5K6	Q8c5k6 mus musculus
12	2365.5	81.6	632	11 Q8VHM5	Q8vhm5 mus musculus
13	2343.5	80.9	636	4 Q9BV64	Q9bv64 homo sapien
14	2227	76.8	601	11 Q99KG1	Q99kg1 mus musculus
15	2023	69.8	410	4 Q81W78	Q81w78 homo sapien
16	1667	57.5	380	11 Q9CT37	Q9ct37 mus musculus

17	1379.5	47.6	707	5 Q9VDI8	Q9vdi8 drosophila
18	1308	45.1	529	5 Q8IN47	Q8in47 drosophila
19	1308	45.1	571	5 Q8IN48	Q8in48 drosophila
20	1300	44.9	529	5 Q95TW4	Q95tw4 drosophila
21	1152.5	39.8	611	5 Q9NLD1	Q9nld1 caenorhabdi
22	939.5	32.4	592	6 Q9XSR3	Q9xsr3 canis famli
23	938	32.4	593	4 Q8NI52	Q8ni52 homo sapien
24	937	32.3	307	11 Q8BL32	Q8bl32 mus musculu
25	936.5	32.3	590	11 Q91WT8	Q91wt8 mus musculu
26	936.5	32.3	590	11 Q8ZE8	Q8ze8 mus musculu
27	895	30.9	586	11 Q924K3	Q924k3 rattus norv
28	893	30.8	594	11 Q923K9	Q923k9 rattus norv
29	890.5	30.7	524	4 Q9NXG3	Q9nxg3 homo sapien
30	882.5	30.5	586	4 Q9NZD3	Q9nzd3 homo sapien
31	882.5	30.5	594	4 Q9NQ94	Q9nq94 homo sapien
32	880.5	30.4	586	4 Q9NQ93	Q9nq93 homo sapien
33	877.5	30.3	586	4 Q9NOX9	Q9nox9 homo sapien
34	876.5	30.2	383	11 Q8CH58	Q8ch58 rattus norv
35	876.5	30.2	405	11 Q8CH57	Q8ch57 rattus norv
36	859.5	29.7	533	4 Q8TBY0	Q8tby0 homo sapien
37	829	28.6	569	4 Q9NOX8	Q9nox8 homo sapien
38	513.5	17.7	495	10 Q9ASP6	Q9asp6 arabidopsis
39	495.5	17.1	521	10 Q23093	Q23093 arabidopsis
40	477	16.5	472	10 Q8W2R4	Q8w2r4 oryza sativ
41	448.5	15.5	431	4 Q9NXC9	Q9nxc9 homo sapien
42	435.5	15.0	471	10 Q9LXJ8	Q9lxj8 arabidopsis
43	386.5	13.3	551	5 Q76494	Q76494 dictyosteli
44	376	13.0	809	10 Q8RWQ1	Q8rwq1 arabidopsis
45	348.5	12.0	353	4 Q81YX4	Q81yx4 homo sapien

ALIGNMENTS

RESULT 1

Q8BGP1	PRELIMINARY;	PRT;	562 AA.		
ID	Q8BGP1				
AC	Q8BGP1;				
DT	01-MAR-2003 (TREMBlrel. 23, Created)				
DT	01-MAR-2003 (TREMBlrel. 23, Last sequence update)				
DT	01-MAR-2003 (TREMBlrel. 23, Last annotation update)				
DE	NSI-associated protein 1.				
DE	Mus musculus (Mouse).				
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J; TISSUE=Body, and Embryo;				
RX	MEDLINE=22354683; PubMed=12466851;				
RA	The FANTOM Consortium,				
RA	"Analysis of the mouse transcriptome based on functional annotation of				
RT	60,770 full-length cDNAs."				
RT	Nature 420:563-573(2002).				
RL	EMBL; AK034845; BAC28852.1; -				
DR	EMBL; AK077588; BAC36880.1; -				
SQ	SEQUENCE 562 AA; 62672 MW; 810BCFA286105377 CRC64;				
Query Match 98.7%; Score 2861; DB 11; Length 562;					
Best Local Similarity 98.0%; Pred. No. 1.5e-191;					
Matches 539; Conservative 3; Mismatches 2; Indels 6; Gaps					
QY	1	MATEHVNGTPEPMDTTS	SAVIHSENFQTL	DAGLPQKVAEKLDIYVAGLVAHS	OLDER
Db	1	MATEHVNGTPEPMDTTS	SAVIHSENFQTL	DAGLPQKVAEKLDIYVAGLVAHS	OLDER
QY	61	AIEALKENEDGALAVLQ	QFQKDSLSHVONKSAFLCGVMKTYRQREKQTKVADS	SGKPD	
Db	61	AIEALKENEDGALAVLQ	QFQKDSLSHVONKSAFLCGVMKTYRQREKQTKVADS	SGKPD	
QY	121	EAKIKALLERTGYTL	DTVTGQRYKYGPPDP	SVYSGQQPSVGTEIFVGKIPDL	FEDELVP

source	1. .2079																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		
--------	----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 28, 2004, 22:58:47 ; Search time 5125 Seconds  
(without alignments)  
16595.339 Million cell updates/sec

Title: US-09-978-242-2  
Perfect score: 2079  
Sequence: 1 GGGCGCGCGCGCGACCGGG.....GTTATTAACTAAAGCTACT 2079

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 2888711 seqs, 2045481386 residues  
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vi:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vi:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htg\_mus:\*
- 34: em\_htg\_pin:\*
- 35: em\_htg\_rod:\*
- 36: em\_htg\_mam:\*
- 37: em\_htg\_vrt:\*
- 38: em\_sy:\*
- 39: em\_htgo\_hum:\*
- 40: em\_htgo\_mus:\*
- 41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2079	100.0	2079	6	AR177851	AR177851 Sequence
2	2012.6	96.8	2221	9	AF155568	AF155568 Homo sapi
3	1883.2	90.6	2208	10	BC041148	BC041148 Mus muscu
4	1848.4	88.9	103819	9	HSJ075C13	AL109618 Human DNA
5	1848.4	88.9	212659	2	AL390737	AL390737 Homo sapi
6	1826.2	87.8	3452	6	BD143848	BD143848 RNA-bind
7	1826.2	87.8	3452	10	AB035725	AB035725 Mus muscu
8	1810.6	87.1	2932	9	AF037448	AF037448 Homo sapi
9	1809.2	87.0	3671	10	BC050079	BC050079 Mus muscu
10	1695.4	81.5	3319	9	BC032643	BC032643 Homo sapi
11	1619.8	77.9	1941	9	BC040844	BC040844 Homo sapi
12	1610.2	77.5	1686	9	AY034483	AY034483 Homo sapi
13	1610.2	77.5	1872	9	AY034481	AY034481 Homo sapi
14	1502.8	72.3	2373	10	AF093821	AF093821 Mus muscu
15	1463.4	70.4	3101	6	AX011753	AX011753 Sequence
16	1390.2	66.9	1767	9	AY034482	AY034482 Homo sapi
17	1329.2	63.9	2232	10	AF408434	AF408434 Mus muscu
18	1089.4	52.4	259969	2	AC118904	AC118904 Rattus no
19	1089.4	52.4	288143	2	AC109106	AC109106 Rattus no
20	935.2	45.0	2371	5	BC046902	BC046902 Danio rer
21	901.8	43.4	222602	2	AC122217	AC122217 Mus muscu
22	895	43.0	1899	10	AF441128	AF441128 Mus muscu
23	895	43.0	2725	10	BC038051	BC038051 Mus muscu
24	884.4	42.5	1899	10	AY184814	AY184814 Rattus no
25	882	42.4	2644	9	AF000364	AF000364 Homo sapi
26	864.6	41.6	2722	9	BC001449	BC001449 Homo sapi
27	852.8	41.0	2471	10	BC004679	BC004679 Mus muscu
28	846.4	40.7	235928	2	AC126148	AC126148 Rattus no
29	846.4	40.7	264178	2	AC098660	AC098660 Rattus no
30	792.4	38.1	934	6	AX013715	AX013715 Sequence
31	763	36.7	196488	2	AC084410	AC084410 Mus muscu
32	763	36.7	205520	2	AC079134	AC079134 Mus muscu
33	705.6	33.9	202374	10	AC122868	AC122868 Mus muscu
34	693.8	33.4	204962	2	AC127932	AC127932 Rattus no
35	693.8	33.4	225370	2	AC129693	AC129693 Rattus no
36	686.6	33.0	215801	2	AC140330	AC140330 Mus muscu
37	642.2	30.9	139996	9	AL161799	AL161799 Human DNA
38	482.8	23.2	2195	10	BC004001	BC004001 Mus muscu
39	445.2	21.4	1296	3	AK116214	AK116214 Ciona int
40	435.2	20.9	263129	2	AC098289	AC098289 Rattus no
41	424.6	20.4	121727	2	AC119124	AC119124 Rattus no
42	424.6	20.4	241150	2	AC105884	AC105884 Rattus no
43	407.2	19.6	265229	2	AC121468	AC121468 Rattus no
44	395.8	19.0	483	6	BD143847	BD143847 RNA-bind
45	378.6	18.2	2606	3	AY058477	AY058477 Drosophil

ALIGNMENTS

RESULT 1	AR177851	2079 bp	DNA	linear	PAT 17-DEC-2001
LOCUS	Sequence 2 from patent US 6313266.				
DEFINITION	AR177851				
ACCESSION	AR177851				
VERSION	AR177851.1	GI:17920206			
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unclassified.				
REFERENCE	1 (bases 1 to 2079)				
AUTHORS	Bandman, O., Yue, H., Corley, N.C. and Shah, P.				
TITLE	Human nucleolin-like protein				
JOURNAL	Patent: US 6313266-A 2 06-NOV-2001;				
FEATURES	Location/Qualifiers				

XX	WPI: 1999-443600/37.	QY	661	TTTCAGGTGAGAGGCTTCTGTTGGCACTGAGATATTTGTTGGGAAAGATCCCAAGAGATCT	720
DR	P-PSDB; AAY24418.	DB	661	TTTCAGGTGAGAGGCTTCTGTTGGCACTGAGATATTTGTTGGGAAAGATCCCAAGAGATCT	720
XX	Nucleic acids encoding human nucleolin-like proteins useful for the	QY	721	ATTGAGGATGAACCTTGTTCATTATTTGAGAAAGCTGGACCTATATGGGATCTTCGTCT	780
PT	diagnosis, prevention and treatment of disorders associated with	DB	721	ATTGAGGATGAACCTTGTTCATTATTTGAGAAAGCTGGACCTATATGGGATCTTCGTCT	780
PT	abnormal nucleolin-like protein expression such as cancers,	QY	781	AATGATGGATCCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT	840
XX	autoimmune disorders and Alzheimer's disease	DB	781	AATGATGGATCCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT	840
PS	Claim 4; Fig 1; 33pp; English.	QY	841	AGAAGCAGCTCAGAGGCTGTTAAACCTGATATAATATCATGAAATTCGTTCTCGAAACA	900
XX	The present sequence encodes human nucleolin-like peptide, designated	DB	841	AGAAGCAGCTCAGAGGCTGTTAAACCTGATATAATATCATGAAATTCGTTCTCGAAACA	900
CC	HNLN. HNLN is the main protein component in the nucleolus of eukaryotic	QY	901	TATTGGTGTCTGATCTCAGTTCGCAACATAGGCTTTTGTGGGCTCTATTCTTAAGAG	960
CC	cells and is an essential part of ribosome biosynthesis and also plays	DB	901	TATTGGTGTCTGATCTCAGTTCGCAACATAGGCTTTTGTGGGCTCTATTCTTAAGAG	960
CC	an important role in importing proteins to the nucleus. HNLN may be used	QY	961	TAAACCAAGGAACAGATCTTCAAGAAATTTAGCAAAAGTAAACAGAGGCTCTTACAGAGCT	1020
CC	in the diagnosis, prevention and treatment of disorders associated with	DB	961	TAAACCAAGGAACAGATCTTCAAGAAATTTAGCAAAAGTAAACAGAGGCTCTTACAGAGCT	1020
CC	abnormal expression of HNLN. For example, it may be used to treat	QY	1021	CATTTTATACCAACCGGATGACAAAGAAAAAAGAGAGGCTTTTGTCTTGAATA	1080
CC	cancers (e.g. melanoma, breast cancer and prostate cancer), autoimmune	DB	1021	CATTTTATACCAACCGGATGACAAAGAAAAAAGAGAGGCTTTTGTCTTGAATA	1080
CC	disorders (e.g. autoimmune haemolytic anaemia and inflammatory bowel	QY	1081	TGAAGATCAAAACAGCTGCCAGGCAAGGCGTGTGTTAATGAGTGGTAAAGTCAAGGT	1140
CC	disease) and, in particular, Alzheimer's disease. It may also be used	DB	1081	TGAAGATCAAAACAGCTGCCAGGCAAGGCGTGTGTTAATGAGTGGTAAAGTCAAGGT	1140
CC	to study the function of the HNLN peptides, the formation (biosynthesis) of	QY	1141	CTGGGGGAATGTTGGAACCTGTTCAATGGGCTGATCTTATAGAAGATCTCTGATCTGAGGT	1200
CC	ribosomes and the intake of proteins into the nucleus.	DB	1141	CTGGGGGAATGTTGGAACCTGTTCAATGGGCTGATCTTATAGAAGATCTCTGATCTGAGGT	1200
XX	Sequence 2079 BP; 639 A; 377 C; 535 G; 528 T; 0 other;	QY	1201	TATGGCAAAAGGTAAAAGTGTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	1260
SQ	Query Match 100.0%; Score 2079; DB 20; Length 2079;	DB	1201	TATGGCAAAAGGTAAAAGTGTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	1260
	Best Local Similarity 100.0%; Pred. No. 0;	QY	1261	TTTAGAAAGGCAATTTAGTCAAGTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	1320
	Matches 2079; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	DB	1261	TTTAGAAAGGCAATTTAGTCAAGTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	1320
QY	1 GGGCGCGCGCGCACCGGAGCGCGCTCGAGGCGAGTGGAACTGGATCGGTTTCTG 60	QY	1321	TGGTTCATTTTATGATGAGCGAGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	1380
DB	1 GGGCGCGCGCGCACCGGAGCGCGCTCGAGGCGAGTGGAACTGGATCGGTTTCTG 60	DB	1321	TGGTTCATTTTATGATGAGCGAGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	1380
QY	61 CCAGCGGCGTGTGCTGCGCGGATTTTACAAAGCTTCACTCGCGCGGACACAGGGA 120	QY	1381	CAAAGACTTGGAGGAGAAATATTTGAAATTTGTTTGGCAAGCCACAGATCAGAAAAG	1440
DB	61 CCAGCGGCGTGTGCTGCGCGGATTTTACAAAGCTTCACTCGCGCGGACACAGGGA 120	DB	1381	CAAAGACTTGGAGGAGAAATATTTGAAATTTGTTTGGCAAGCCACAGATCAGAAAAG	1440
QY	121 GCAGCGAGCAGCGGTTTCCGCAACCGGATACCATCGACAGGATTTCTCGCGCTCAGCC 180	QY	1441	GAAAGAAAGAAAGCTCAGAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCA	1500
DB	121 GCAGCGAGCAGCGGTTTCCGCAACCGGATACCATCGACAGGATTTCTCGCGCTCAGCC 180	DB	1441	GAAAGAAAGAAAGCTCAGAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCA	1500
QY	181 CAACGGGAGATCTCTGAAACATGGCTACAGAACATGTTAATGAAATGGTACTGAAGA 240	QY	1501	TTATGGTCCACCTCATATGCTCCCAACAGAGGTCGAGGCGCTGAGGAGTGAAGGTGG	1560
DB	181 CAACGGGAGATCTCTGAAACATGGCTACAGAACATGTTAATGAAATGGTACTGAAGA 240	DB	1501	TTATGGTCCACCTCATATGCTCCCAACAGAGGTCGAGGCGCTGAGGAGTGAAGGTGG	1560
QY	241 GCCCATGGATCTACTTCTGAGTTATCCATTGAGAAATTTTCAGACATTTGCTTGATGC 300	QY	1561	TTATGGATATCTCCAGATTTATGATATGAAGATTTATGATTTATGATTTATGATTTATGA	1620
DB	241 GCCCATGGATCTACTTCTGAGTTATCCATTGAGAAATTTTCAGACATTTGCTTGATGC 300	DB	1561	TTATGGATATCTCCAGATTTATGATATGAAGATTTATGATTTATGATTTATGATTTATGA	1620
QY	301 TGGTTTACCAAGAAAGTGTCTGAAACATAGATGAAATTTACGTTGAGGCGCTAGTTGC 360	QY	1621	TTACCATTAATCTGCTGGTGGATATGAAGATCCATCTATGTTTATGAAGATTTTCAAGT	1680
DB	301 TGGTTTACCAAGAAAGTGTCTGAAACATAGATGAAATTTACGTTGAGGCGCTAGTTGC 360	DB	1621	TTACCATTAATCTGCTGGTGGATATGAAGATCCATCTATGTTTATGAAGATTTTCAAGT	1680
QY	361 ACATAGTGAATTTAGATGAAGAGCTATTGAAAGCTTTTAAAGAAATTTCAATGAAGCGTGC 420	QY	1681	TGGAGCTAGAGGAGGCGTGTAGAGGAGCAAGGCGTGTCTCCATCCAGAGGTCTGTTG	1740
DB	361 ACATAGTGAATTTAGATGAAGAGCTATTGAAAGCTTTTAAAGAAATTTCAATGAAGCGTGC 420	DB	1681	TGGAGCTAGAGGAGGCGTGTAGAGGAGCAAGGCGTGTCTCCATCCAGAGGTCTGTTG	1740
QY	421 ATTGGCAGTTCTTCAACAGTTTAAAGACAGTGTCTCTCATGTTTCAAGAAAGTGC 480	QY	1741	GGCTGCTCTCTCCCGCGGTAGAGCGGCTTTATTTCAAGAGAGGAGGTCTCTGATCAGCAAG	1800
DB	421 ATTGGCAGTTCTTCAACAGTTTAAAGACAGTGTCTCTCATGTTTCAAGAAAGTGC 480	DB			
QY	481 CTTTATGTCGAGTTCATGAAGCTTACAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540	QY			
DB	481 CTTTATGTCGAGTTCATGAAGCTTACAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540	DB			
QY	541 AGATTCTAGTAAAGGACAGATGAGGCAAAATTAAGGCACTCTTTGGAAGAAACAGGCTA 600	QY			
DB	541 AGATTCTAGTAAAGGACAGATGAGGCAAAATTAAGGCACTCTTTGGAAGAAACAGGCTA 600	DB			
QY	601 CACACTTGATGTGACCACTGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660	QY			
DB	601 CACACTTGATGTGACCACTGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660	DB			



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 28, 2004, 22:57:52 ; Search time 411 Seconds  
(without alignments)  
13654.822 Million cell updates/sec

Title: US-09-978-242-2

Perfect score: 2079

Sequence: 1 GGGCGCGCGCGCGCACCGG.....GTTATTAACTAAAGCTACT 2079

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_19Jun03:\*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*
- 25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2079	100.0	2079	20	Human nucleolin-li
2	2079	100.0	2079	20	Human cDNA encodin
3	1844	88.7	4231	21	Human prostate can
4	1826.2	87.8	3452	24	Nucleotide sequenc
5	1777.8	85.5	2680	21	Human RNA-associat
6	1731	83.3	2407	23	DNA encoding novel
7	1463.4	70.4	3101	20	Human prostate tum
8	882	42.4	2673	21	Human secreted pro

c	9	882	42.4	3159	25	ACA03964	cdNA downregulated
10	871.6	41.9	2671	24	ABK35808	cdNA sequence #199	
11	839.8	40.4	2714	22	ABL26717	Human breast cancer	
12	833.6	40.1	2700	22	AH33268	Human colon cancer	
13	792.4	38.1	934	20	AAZ42164	Human normal bladd	
14	730	35.1	1024	21	AAZ98058	Human colon cancer	
15	498.8	24.0	612	21	AAZ80570	Human colon cancer	
16	487	23.4	1907	24	AAK35125	Human cDNA encodin	
17	482.8	23.2	560	24	ABQ58093	Human colon cancer	
18	425	20.4	464	22	ABA51954	Human foetal liver	
19	425	20.4	464	22	ABA21772	Probe #238 for gen	
20	425	20.4	464	22	AAK00240	Human brain expres	
21	425	20.4	464	22	AAK25682	Human bone marrow	
22	425	20.4	464	22	AAI10309	Probe #242 for gen	
23	425	20.4	464	22	AAI31562	Probe #248 used to	
24	425	20.4	464	22	AAI00245	Probe #236 used to	
25	425	20.4	464	24	ABS25267	Human liver single	
26	425	20.4	464	24	ABS00260	Human genome-deriv	
27	395.8	19.0	483	24	ABL59087	Nucleotide sequenc	
28	371.8	17.9	384	19	AAV54582	Human secretory pr	
29	371.8	17.9	384	20	AAZ25601	Human secreted pro	
30	354.6	17.1	747	24	ABS77417	Frog embryonic gen	
31	287.8	13.8	591	22	AAH35088	Human colon cancer	
32	275.2	13.2	700	24	ABS77066	Frog embryonic gen	
33	272.8	13.1	382	22	ABA46733	Human breast cell	
34	272.8	13.1	382	22	ABA64613	Human foetal liver	
35	272.8	13.1	382	22	ABA31740	Probe #10206 for g	
36	272.8	13.1	382	22	AAK13051	Human brain expres	
37	272.8	13.1	382	22	AAK38781	Human bone marrow	
38	272.8	13.1	382	22	AAI19587	Probe #9520 for ge	
39	272.8	13.1	382	22	AAI44783	Probe #13469 used	
40	272.8	13.1	382	22	AAI05307	Probe #5298 used t	
41	272.8	13.1	382	23	ABS38364	Human liver single	
42	272.8	13.1	382	24	ABS12860	Human genome-deriv	
43	262.2	12.6	342	14	AAQ59054	Human brain Expres	
44	261.6	12.6	342	14	AAQ39642	Expressed Sequence	
45	261	12.6	307	20	AAV86918	EST clone BK260.	

#### ALIGNMENTS

RESULT 1  
AAZ90183  
ID AAZ90183 standard; cdNA; 2079 BP.  
XX AAZ90183;  
AC AAZ90183;  
DT 21-SEP-1999 (first entry)  
XX Human nucleolin-like protein encoding cdNA.  
XX Human; nucleolin-like protein; HNL; cancer; melanoma; breast cancer;  
KW prostate cancer; autoimmune disorder; autoimmune haemolytic anaemia;  
KW inflammatory bowel disease; Alzheimer's disease; ss.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
FH Key 203..1840  
FT CDS /\*tag= a  
FT FT  
XX US5932475-A.  
PN  
XX  
PD 03-AUG-1999.  
PF 12-DEC-1997; 97US-0990114.  
XX  
PR 12-DEC-1997; 97US-0990114.  
XX  
PA (INCY-) INCYTE PHARM INC.  
XX Bandman O, Corley NC, Shah P, Yue H;



Best Local Similarity 100.0%; Pred. No. 0; Matches 2079; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
Qy	1	GGGCGCGCGCGCACCGGAGCGCGCTCGAGCGAGTGGAACTGATCGGTTTCTG	60						
Ds	1	GGGCGCGCGCGCACCGGAGCGCGCTCGAGCGAGTGGAACTGATCGGTTTCTG	60						
Qy	61	CCAGCGCGTGAGCTTCCGCGCGGATTTTACACAGCTCCACTCGCGCGGACACAGGGA	120						
Ds	61	CCAGCGCGTGAGCTTCCGCGCGGATTTTACACAGCTCCACTCGCGCGGACACAGGGA	120						
Qy	121	GCAGCGAGCAGCGTTTCCGCAACCCGATACCATCGGACAGGATTTCTCGCGCTCAGCC	180						
Ds	121	GCAGCGAGCAGCGTTTCCGCAACCCGATACCATCGGACAGGATTTCTCGCGCTCAGCC	180						
Qy	181	CAACGGGAGATCTCTGGAAACATGGCTACAGACATGTTAATGGAAATGGTACTGAAGA	240						
Ds	181	CAACGGGAGATCTCTGGAAACATGGCTACAGACATGTTAATGGAAATGGTACTGAAGA	240						
Qy	241	GCCCATGGATACTTCTGCGATTTATCCATTAGAAATTTTTCAGACATTTGCTTGATGC	300						
Ds	241	GCCCATGGATACTTCTGCGATTTATCCATTAGAAATTTTTCAGACATTTGCTTGATGC	300						
Qy	301	TGGTTTACCAGAAAGTTGCTGAAACCTAGATGAATTTTACGTTGCGGCTAGTTGC	360						
Ds	301	TGGTTTACCAGAAAGTTGCTGAAACCTAGATGAATTTTACGTTGCGGCTAGTTGC	360						
Qy	361	ACATAGTGAATTTAGATGAAGAGCTTATGAAGCTTTAAAGAAATTTCAATGAAGCGGTGC	420						
Ds	361	ACATAGTGAATTTAGATGAAGAGCTTATGAAGCTTTAAAGAAATTTCAATGAAGCGGTGC	420						
Qy	421	ATTGGCAGTCTCTCAACAGTTTAAAGACAGTGATCTCTCATGTTTCAAGCAAAAGTGC	480						
Ds	421	ATTGGCAGTCTCTCAACAGTTTAAAGACAGTGATCTCTCATGTTTCAAGCAAAAGTGC	480						
Qy	481	CTTTTATGTGGAGTCATGAAGACTTACAGCAGAGAGAAAAACAAGGACCAAAAGTAGC	540						
Ds	481	CTTTTATGTGGAGTCATGAAGACTTACAGCAGAGAGAAAAACAAGGACCAAAAGTAGC	540						
Qy	541	AGATTCTAGTAAAGACAGATGAGGCAAAATTTAAGCACTCTTGGAAAGACAGGCTA	600						
Ds	541	AGATTCTAGTAAAGACAGATGAGGCAAAATTTAAGCACTCTTGGAAAGACAGGCTA	600						
Qy	601	CACACTGTGACCTGACAGAGAGAGTATGGAGGACCACTCCAGATTCCGTTTA	660						
Ds	601	CACACTGTGACCTGACAGAGAGAGTATGGAGGACCACTCCAGATTCCGTTTA	660						
Qy	661	TTCAGGTGACAGCGCTTCTGTGGCAGTGAATATTTGGGAAAGATCCCAAGAGATCT	720						
Ds	661	TTCAGGTGACAGCGCTTCTGTGGCAGTGAATATTTGGGAAAGATCCCAAGAGATCT	720						
Qy	721	ATTTGAGGATGAATTTGTTTCAATTTTGGAAAGCTGGACCTATATGGGATCTTCTCT	780						
Ds	721	ATTTGAGGATGAATTTGTTTCAATTTTGGAAAGCTGGACCTATATGGGATCTTCTCTCT	780						
Qy	781	AATGATGATCCACTCAGTGTCTCAATAGAGGTTATGCGTTTGTACATAA	840						
Ds	781	AATGATGATCCACTCAGTGTCTCAATAGAGGTTATGCGTTTGTACATAA	840						
Qy	841	AGAAGCAGCTCAGGAGGCTTTAACTGTATAAATCATGAATTCGTTCTGGAAACA	900						
Ds	841	AGAAGCAGCTCAGGAGGCTTTAACTGTATAAATCATGAATTCGTTCTGGAAACA	900						
Qy	901	TATTGGTGTGATCTCAGTTGCCAATAGGCTTTTGTGGGCTCTATCTTAAGAG	960						
Ds	901	TATTGGTGTGATCTCAGTTGCCAATAGGCTTTTGTGGGCTCTATCTTAAGAG	960						
Qy	961	TAAACCAAGGAACAGATTTCTTGAAGATTTAGCAAGATTAACAGAGGCTTTACAGAGT	1020						
Ds	961	TAAACCAAGGAACAGATTTCTTGAAGATTTAGCAAGATTAACAGAGGCTTTACAGAGT	1020						
Qy	1021	CATTTTATACCACCGGATGACAGAAAAAACAAGAGGCTTTTCTTTTGAATA	1080						

Db	1021	CATTTTATACCACCGGATGACAGAAAAAACAAGAGGCTTTTCTTTTGAATA	1080
Qy	1081	TGAAGATCAAAACAGCTGCCAGGCAAGGCGTGTAGTTTAACTAGTGGTAAAGTCAAGGT	1140
Ds	1081	TGAAGATCAAAACAGCTGCCAGGCAAGGCGTGTAGTTTAACTAGTGGTAAAGTCAAGGT	1140
Qy	1141	CTGGGGAATGTTGGAACTGTTGAATGGGCTGATCTATAGAAAGATCTCTGAGGT	1200
Ds	1141	CTGGGGAATGTTGGAACTGTTGAATGGGCTGATCTATAGAAAGATCTCTGAGGT	1200
Qy	1201	TATGGCAAGGTTAAAGTGTCTGTACGCAACCTTGGCAATCTGTACAGAGAGAT	1260
Ds	1201	TATGGCAAGGTTAAAGTGTCTGTACGCAACCTTGGCAATCTGTACAGAGAGAT	1260
Qy	1261	TTTATAGAAAGGCAATTTAGTCAGTTTGGGAAAATGGAAACGAGTGAAGAAATTA	1320
Ds	1261	TTTATAGAAAGGCAATTTAGTCAGTTTGGGAAAATGGAAACGAGTGAAGAAATTA	1320
Qy	1321	TGCGTTTCAATTTTGTATGAGCGAGATGTTGCTGCTCAAGGCTATGGAGAAATGAATGG	1380
Ds	1321	TGCGTTTCAATTTTGTATGAGCGAGATGTTGCTGCTCAAGGCTATGGAGAAATGAATGG	1380
Qy	1381	CAAGACTTGGAGGGGAGAAATTTGAAATTTGTTTGGCAAGCCACAGATCAGAAAAG	1440
Ds	1381	CAAGACTTGGAGGGGAGAAATTTGAAATTTGTTTGGCAAGCCACAGATCAGAAAAG	1440
Qy	1441	GAAGAAAGAAAGCTCAGAGGCAAGCAGCAAAATTTGTTTGGCAAGCCACAGATCAGAAAAG	1500
Ds	1441	GAAGAAAGAAAGCTCAGAGGCAAGCAGCAAAATTTGTTTGGCAAGCCACAGATCAGAAAAG	1500
Qy	1501	TTATGTTTCCATCTATATGCTCCCTCCAAAGAGGTCGAGGCGGTGAGAGGTGAGAGTGG	1560
Ds	1501	TTATGTTTCCATCTATATGCTCCCTCCAAAGAGGTCGAGGCGGTGAGAGGTGAGAGTGG	1560
Qy	1561	TTATGATATCTCCAGATTTATGATATGAAGATTTATGATTTATGTTTATGA	1620
Ds	1561	TTATGATATCTCCAGATTTATGATATGAAGATTTATGATTTATGTTTATGA	1620
Qy	1621	TTACCAATACTATCGTGGTGATATGAAGATCCATCTACTATGTTTATGAAGATTTTCAAGT	1680
Ds	1621	TTACCAATACTATCGTGGTGATATGAAGATCCATCTACTATGTTTATGAAGATTTTCAAGT	1680
Qy	1681	TGGAGCTTAGAGGAAGGGTGGTAGAGGACGAGGGGTGCTTCCATCCAGAGGTCTGCG	1740
Ds	1681	TGGAGCTTAGAGGAAGGGTGGTAGAGGACGAGGGGTGCTTCCATCCAGAGGTCTGCG	1740
Qy	1741	GGCTGCTCTCCCGGCTAGAGCGGTTATTCACAGAGGAGGTCCTCGATCAGCAAG	1800
Ds	1741	GGCTGCTCTCCCGGCTAGAGCGGTTATTCACAGAGGAGGTCCTCGATCAGCAAG	1800
Qy	1801	AGGCGTTTCGAGCAGGGAAGGGGTGAGGCGGCTGCTGACCTGTTTACAATGAAGACTGA	1860
Ds	1801	AGGCGTTTCGAGCAGGGAAGGGGTGAGGCGGCTGCTGACCTGTTTACAATGAAGACTGA	1860
Qy	1861	CTTGCTATGTGGGATTTACACAGAGCTTCAGTGGAGTAAATGGTAAGGAAATCAAGCAA	1920
Ds	1861	CTTGCTATGTGGGATTTACACAGAGCTTCAGTGGAGTAAATGGTAAGGAAATCAAGCAA	1920
Qy	1921	CTTTAAATATGTCGGCTGTATAGGACATTTCTTATGGAGAGACCTTCTTATGAAGAT	1980
Ds	1921	CTTTAAATATGTCGGCTGTATAGGACATTTCTTATGGAGAGACCTTCTTATGAAGAT	1980
Qy	1981	CATGGAATCAAAATACGGGACATTTGAACATACTTTGACCTTTGATATGAATTTCTTTAAC	2040
Ds	1981	CATGGAATCAAAATACGGGACATTTGAACATACTTTGACCTTTGATATGAATTTCTTTAAC	2040
Qy	2041	AAATTTCTCTGCGAGTCAAGTATTAACTTAAAGTACT 2079	
Ds	2041	AAATTTCTCTGCGAGTCAAGTATTAACTTAAAGTACT 2079	

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 00:15:58 ; Search time 109 Seconds  
(without alignments)  
8418.675 Million cell updates/sec

Title: US-09-978-242-2

Perfect score: 2079

Sequence: 1 GGGCGCGCGCGCGCGCGCGG.....GTTATTAACTAAAGTACT 2079

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:\*

1: /cgn2\_6/ptodata/2/ina/5A COMB.seq:\*

2: /cgn2\_6/ptodata/2/ina/5B COMB.seq:\*

3: /cgn2\_6/ptodata/2/ina/6A COMB.seq:\*

4: /cgn2\_6/ptodata/2/ina/6B COMB.seq:\*

5: /cgn2\_6/ptodata/2/ina/PTCUS COMB.seq:\*

6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2079	100.0	2079	2	US-08-990-114-2
2	2079	100.0	2079	4	US-09-241-333-2
3	498.8	24.0	612	3	US-09-328-111-654
4	371.8	17.9	384	2	US-09-014-969-3
5	65.6	3.2	7218	1	US-08-232-463-14
6	48.4	2.3	2361	4	US-09-152-060-28
7	47.8	2.3	18596	3	US-09-318-448-11
8	47.8	2.3	18597	4	US-09-962-665-8
9	47.2	2.3	2277	1	US-08-676-967-2
10	47.2	2.3	2277	1	US-08-676-974-2
11	47.2	2.3	2277	2	US-09-098-487-2
12	46.4	2.2	1894	2	US-08-935-450-7
13	44.8	2.2	309	1	US-08-209-747-3
14	44.8	2.2	309	1	US-08-458-298-3
15	44.2	2.1	1505	1	US-07-915-246-1
16	44.2	2.1	4673	1	US-07-638-431-1
17	44.2	2.1	4673	5	PCT-US92-00018-1
18	43.6	2.1	10640	4	US-09-417-4850-5
19	43.2	2.1	2042	2	US-08-911-434A-2
20	42.6	2.0	527	3	US-09-347-833-5
21	42.4	2.0	5340	2	US-09-627-122-21
22	42.4	2.0	8920	2	US-08-446-855A-1
23	42.4	2.0	8920	3	US-09-150-741-1
24	42.2	2.0	2793	1	US-08-209-747-1
25	42.2	2.0	2793	1	US-08-458-298-1
26	42.2	2.0	1664976	4	US-08-916-421B-1
27	41.8	2.0	5340	4	US-09-627-122-21

Sequence 4, Appli  
Sequence 4, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 20, Appli  
Sequence 20, Appli  
Sequence 7, Appli  
Sequence 19, Appli  
Sequence 3, Appli  
Sequence 2161, Ap  
Sequence 1, Appli  
Sequence 2, Appli  
Sequence 4, Appli

#### ALIGNMENTS

RESULT 1  
US-08-990-114-2  
; Sequence 2, Application US/08990114  
; Patent No. 5932475  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Yue, Henry  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Shah, Purvi  
; TITLE OF INVENTION: HUMAN NUCLEOLIN-LIKE PROTEIN  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/990,114  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0451 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2079 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: TLYNOT05  
; CLONE: 2809795  
; US-08-990-114-2

Query Match 100.0%; Score 2079; DB 2; Length 2079;

```

;
; TELEFAX: 650-845-4166
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 2079 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; IMMEDIATE SOURCE:
;   LIBRARY: TLYMN0705
;   CLONE: 2809795
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
us-09-978-242-2

Query Match      100.0%; Score 2079; DB 9; Length 2079;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2079; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGCGCGCGCGCGCGCGCGCGCGCTCGGAGCGAGTGGAACTGGATCGGTTTCTG 60
Db 1 GGGCGCGCGCGCGCGCGCGCGCGCTCGGAGCGAGTGGAACTGGATCGGTTTCTG 60

Qy 61 CCAGCGCGGTGAGCTTCGGCCGGGCAATTTTACAACAGCTCCACTCGCGCGGACACAGGA 120
Db 61 CCAGCGCGGTGAGCTTCGGCCGGGCAATTTTACAACAGCTCCACTCGCGCGGACACAGGA 120

Qy 121 GCAGCGACGCGCTTCGCCCAACCCGATACCATCGGACAGGATTTCTCGCCTCAGCC 180
Db 121 GCAGCGACGCGCTTCGCCCAACCCGATACCATCGGACAGGATTTCTCGCCTCAGCC 180

Qy 181 CAACGGGAGATCTCTGGAACATGGCTACAGAACATGTTAATGGAAATGGTACTGAAGA 240
Db 181 CAACGGGAGATCTCTGGAACATGGCTACAGAACATGTTAATGGAAATGGTACTGAAGA 240

Qy 241 GCCCATGATATCTCTTCGAGTTATCCATTCAAGAAATTTTCAGACATTCGTTGATGC 300
Db 241 GCCCATGATATCTCTTCGAGTTATCCATTCAAGAAATTTTCAGACATTCGTTGATGC 300

Qy 301 TGGTTTACCAGAAAGTTGCTGAAACACTAGATGAATTTACGTTGCGAGGCTAGTTGC 360
Db 301 TGGTTTACCAGAAAGTTGCTGAAACACTAGATGAATTTACGTTGCGAGGCTAGTTGC 360

Qy 361 ACATAGTGAATTTAGATGAAGAGCTATTGAAGCTTTTAAAGAAATTCATGAAGACGGTGC 420
Db 361 ACATAGTGAATTTAGATGAAGAGCTATTGAAGCTTTTAAAGAAATTCATGAAGACGGTGC 420

Qy 421 ATTGGCAGTTCTTCAACAGTTTAAAGACAGTGTCTCTCATGTTTCAAGAAACAAAGTGC 480
Db 421 ATTGGCAGTTCTTCAACAGTTTAAAGACAGTGTCTCTCATGTTTCAAGAAACAAAGTGC 480

Qy 481 CTTTTTATGTGGAGTCATGAAGCTTTACAGCAGAGAGAAACAAAGGACCAAGTAGC 540
Db 481 CTTTTTATGTGGAGTCATGAAGCTTTACAGCAGAGAGAAACAAAGGACCAAGTAGC 540

Qy 541 AGATTCTAGTAAAGACACAGATGAGGCAAAATTAAGGCACCTCTTGGAAAGAACAGGCTA 600
Db 541 AGATTCTAGTAAAGACACAGATGAGGCAAAATTAAGGCACCTCTTGGAAAGAACAGGCTA 600

Qy 601 CACACTGATGTGACCACTGGACAGAGAAAGTATGGAGGACCACTTCAGATTCGGTTTA 660
Db 601 CACACTGATGTGACCACTGGACAGAGAAAGTATGGAGGACCACTTCAGATTCGGTTTA 660

Qy 661 TTCAGTTCAGGACCTCTGTTGGCAGCTGAGATATTTCTGGGAAAGATCCCAAGAGATCT 720
Db 661 TTCAGTTCAGGACCTCTGTTGGCAGCTGAGATATTTCTGGGAAAGATCCCAAGAGATCT 720

Qy 721 ATTTGAGGATGAATCTGTTTCCATTATTTGAGAAAGCTGACCTATATGGGATCTTCGTCT 780
Db 721 ATTTGAGGATGAATCTGTTTCCATTATTTGAGAAAGCTGACCTATATGGGATCTTCGTCT 780

Qy 781 AATGATGATCCACTCAGCTGCTCAATAGAGGTTATCGGTTTGTCACTTTTGTACAAA 840
Db 781 AATGATGATCCACTCAGCTGCTCAATAGAGGTTATCGGTTTGTCACTTTTGTACAAA 840

841 AGAAGCAGCTCAGGAGGCTGTTAAACTGTATATATCATGAAATTCGTTCTGGAACA 900
841 AGAAGCAGCTCAGGAGGCTGTTAAACTGTATATATATCATGAAATTCGTTCTGGAACA 900

901 TATTGTGCTGCTCATCTCAGTTGCGCAACAAATAGGCTTTTGTGGGCTCTATTCTTAAGAG 960
901 TATTGTGCTGCTCATCTCAGTTGCGCAACAAATAGGCTTTTGTGGGCTCTATTCTTAAGAG 960

961 TAAACCAAGGACAGATTTCTTGAAGAAATTTAGCAAGTAACAGAGGGTCTTACAGACGT 1020
961 TAAACCAAGGACAGATTTCTTGAAGAAATTTAGCAAGTAACAGAGGGTCTTACAGACGT 1020

1021 CATTTTATATACCACCGGATGCAAGAAATTTTGAAGGCTTTTGTCTTCTTGAATA 1080
1021 CATTTTATATACCACCGGATGCAAGAAATTTTGAAGGCTTTTGTCTTCTTGAATA 1080

1081 TGAAGATCAAAAACAGCTGCCAGCAAGGCTAGTGTAAAGTGTAAAGTCAAGGT 1140
1081 TGAAGATCAAAAACAGCTGCCAGCAAGGCTAGTGTAAAGTGTAAAGTCAAGGT 1140

1141 CTGGGGGAATGTTGGAACTGTTGAAATGGCTGATCTCTATAGAAAGATCTCTGATCCTGAGGT 1200
1141 CTGGGGGAATGTTGGAACTGTTGAAATGGCTGATCTCTATAGAAAGATCTCTGATCCTGAGGT 1200

1201 TATGGCAAAAGGTAAAGTGTCTGTTGTACGCAACCTTGCCTATAGAAAGATCTCTGATCCTGAGGT 1260
1201 TATGGCAAAAGGTAAAGTGTCTGTTGTACGCAACCTTGCCTATAGAAAGATCTCTGATCCTGAGGT 1260

1261 TTTAGAAAAGGCAATTTAGTCAGTTTGGGAAACTGGAAACGAGTGAAAGTATAAGATTA 1320
1261 TTTAGAAAAGGCAATTTAGTCAGTTTGGGAAACTGGAAACGAGTGAAAGTATAAGATTA 1320

1321 TGGTTTCATTCATTTTGTAGTGAGCGAGATGCTGCTCAAGGCTATGGAAGAAATGAATGG 1380
1321 TGGTTTCATTCATTTTGTAGTGAGCGAGATGCTGCTCAAGGCTATGGAAGAAATGAATGG 1380

1381 CAAAGACTTGGAGGAGAAATAATTGAAATTTGTTTGGCAAGCCACCAGATCAGAAAG 1440
1381 CAAAGACTTGGAGGAGAAATAATTGAAATTTGTTTGGCAAGCCACCAGATCAGAAAG 1440

1441 GAAAGAAAGAAAGCTCAGAGGCAAGCAGCAAAAATCAATGTATGACGATTACTACTA 1500
1441 GAAAGAAAGAAAGCTCAGAGGCAAGCAGCAAAAATCAATGTATGACGATTACTACTA 1500

1501 TTTATGTCACCTCATATGCCCTCCCAACAGAGTCCAGGCGCTGAGGTAGAGGTGG 1560
1501 TTTATGTCACCTCATATGCCCTCCCAACAGAGTCCAGGCGCTGAGGTAGAGGTGG 1560

1561 TTATGGATATCTCCAGATTTATTTATGGATATGAAGATTTATTTATGTTTATGA 1620
1561 TTATGGATATCTCCAGATTTATTTATGGATATGAAGATTTATTTATGTTTATGA 1620

1621 TTACCAATACTATCTGTTGGATATGAAGTCCATATCTATGTTTATGAAGATTTTCAAGT 1680
1621 TTACCAATACTATCTGTTGGATATGAAGTCCATATCTATGTTTATGAAGATTTTCAAGT 1680

1681 TGGAGCTAGAGGAGGCTGTTAGAGGACCAAGGGTGTCTCCATCCAGAGGTCTGTTG 1740
1681 TGGAGCTAGAGGAGGCTGTTAGAGGACCAAGGGTGTCTCCATCCAGAGGTCTGTTG 1740

1741 GGCTGCTCTCTCCCGGGTAGAGCCGGTATTTCACAGAGGAGGCTCTGGATCAGCAAG 1800
1741 GGCTGCTCTCTCCCGGGTAGAGCCGGTATTTCACAGAGGAGGCTCTGGATCAGCAAG 1800

1801 AGCGTTTCGAGCAGGAGGCTGAGGCGGCTCTGACCTGTTTACATGAAGACTGA 1860
1801 AGCGTTTCGAGCAGGAGGCTGAGGCGGCTCTGACCTGTTTACATGAAGACTGA 1860

1861 CTTGCTATGTGGATTTACACAGAGCTTGCAGTGGAGTAAATGTAAGGAAATCAAGCAA 1920
1861 CTTGCTATGTGGATTTACACAGAGCTTGCAGTGGAGTAAATGTAAGGAAATCAAGCAA 1920
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 01:57:54 ; Search time 532 Seconds  
(without alignments)

13976.470 Million cell updates/sec

Title: US-09-978-242-2

Perfect score: 2079

Sequence: 1 GGGCGCGCGCGCGCACCGG.....GTTATTAACTAAAGTACT 2079

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2356869 seqs, 1788235258 residues

Total number of hits satisfying chosen parameters: 4713738

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2079	100.0	2079	9	US-09-978-242-2
2	1844	88.7	4231	10	US-09-925-300-475
3	1826.2	87.8	3452	10	US-09-821-687-3
4	1813.4	87.2	2369	15	US-10-106-698-1042
5	882	42.4	2673	11	US-09-374-046A-127
6	882	42.4	3159	14	US-10-071-766-85
7	871.6	41.9	2671	11	US-09-822-846-199
8	839.8	40.4	2073	15	US-10-198-846-13811
9	833.6	40.1	2701	15	US-10-106-698-334
10	730	35.1	1024	9	US-09-925-299-68
11	730	35.1	1024	11	US-09-925-299-68
12	498.8	24.0	612	10	US-09-879-536-654
13	487	23.4	1907	9	US-09-822-849A-263
14	434.8	20.9	1176	15	US-10-198-846-9841
15	425	20.4	464	9	US-09-864-761-238

16	414.4	19.9	418	10	US-09-796-692-7702	Sequence 7702, Ap
17	414.4	19.9	418	12	US-10-057-475B-7702	Sequence 7702, Ap
18	414.4	19.9	418	12	US-10-154-884B-7702	Sequence 7702, Ap
19	414.4	19.9	418	15	US-10-040-862-7702	Sequence 7702, Ap
20	395.8	19.0	483	10	US-09-821-687-1	Sequence 1, Appli
21	391	18.8	403	11	US-09-918-995-8445	Sequence 8445, Ap
c	367	17.7	2715	15	US-10-106-698-1041	Sequence 1041, Ap
23	354.6	17.1	747	9	US-09-910-943-643	Sequence 643, App
24	287.8	13.8	591	15	US-10-106-698-2180	Sequence 2180, Ap
25	275.2	13.2	700	9	US-09-910-943-301	Sequence 301, App
26	272.8	13.1	382	9	US-09-864-761-17060	Sequence 17060, A
27	257.6	12.4	2022	13	US-09-949-029-99	Sequence 99, Appl
28	253.2	12.2	1992	14	US-10-098-841-310	Sequence 310, App
29	249.2	12.0	408	11	US-09-918-995-18621	Sequence 18621, A
30	248.4	11.9	500	11	US-09-918-995-24031	Sequence 24031, A
31	215	10.3	387	15	US-10-198-846-5910	Sequence 5910, Ap
32	212	10.2	368	10	US-09-960-352-9809	Sequence 9809, Ap
33	192.6	9.3	405	11	US-09-918-995-33148	Sequence 33148, A
34	180.4	8.7	429	10	US-09-738-973-414	Sequence 414, App
35	180.4	8.7	429	10	US-09-854-133-414	Sequence 414, App
36	180.4	8.7	429	15	US-10-144-649A-414	Sequence 414, App
37	165.2	7.9	480	15	US-10-198-846-7972	Sequence 7972, Ap
38	155.2	7.5	608	9	US-09-764-887-48	Sequence 48, Appl
39	155.2	7.5	608	15	US-10-073-961-48	Sequence 48, Appl
40	153.2	7.4	589	13	US-10-029-386-12112	Sequence 12112, A
41	141.6	6.8	286	10	US-09-867-701-12	Sequence 12, Appl
42	120.2	5.8	761	9	US-09-910-943-415	Sequence 415, App
43	110.2	5.3	220	13	US-10-029-386-25812	Sequence 25812, A
c	90.6	4.4	626	13	US-10-027-632-291227	Sequence 291227, A
44	90.6	4.4	626	13	US-10-027-632-291228	Sequence 291228, A
45	90.6	4.4	626	13	US-10-027-632-291228	Sequence 291228, A

#### ALIGNMENTS

#### RESULT 1

US-09-978-242-2  
; Sequence 2, Application US/09978242  
; Patent No. US20020098566A1  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; Yue, Henry  
; Corley, Neil C.  
; Shah, Purvi  
; TITLE OF INVENTION: HUMAN NUCLEOLIN-LIKE PROTEIN  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/978,242  
; FILING DATE: 15-Oct-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/241,333  
; FILING DATE: <Unknown>  
; APPLICATION NUMBER: 08/990,114  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0451 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555

REFERENCE AUTHORS	3	Shibata, K., Itoh, M., Aizawa, K., Nagao, K., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsu, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
TITLE		RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL MEDLINE PUBMED		Genome Res. 10 (11), 1757-1771 (2000)
REFERENCE AUTHORS	4	20530913 11076861
REFERENCE AUTHORS	5	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G., Quackenbush, J., Schriml, L.M., Stauber, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Balderelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D.A., Kaniwa, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, C., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schombach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyooka, K., Wang, K.H., Weitz, C., Wittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S., and Hayashizaki, Y.
TITLE		Functional annotation of a full-length mouse cDNA collection
JOURNAL MEDLINE PUBMED		Nature 409 (6821), 685-690 (2001)
REFERENCE AUTHORS	5	21085660 11217851
REFERENCE AUTHORS	6	(bases 1 to 3838)
TITLE		The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
JOURNAL MEDLINE PUBMED		Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
REFERENCE AUTHORS	6	Nature 420, 563-573 (2002)
REFERENCE AUTHORS	7	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hangaki, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.
TITLE		Direct Submission
JOURNAL		Submitted (16-JUL-2001) Yoshinori Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@ac.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT		cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.go.jp/ URL: http://fantom.gsc.riken.go.jp/ Location/Qualifiers
FEATURES		

source	1..3838	/organism="Mus musculus"
		/mol_type="mRNA"
		/strain="C57BL/6J"
		/db_xref="PANTOM DB:9430046J23"
		/db_xref="taxon:10090"
		/clone="9430046J23"
		/issue_type="embryonic body between diaphragm region and neck"
		/clone_lib="RIKEN full-length enriched mouse cDNA library"
		/dev_stage="12 days embryo"
CDS	167..1855	/note="unmated protein product; NSI-associated protein 1 (MGDI) [GI:1891690, GI:AB035725, evidence: BLASTN, 99%, match=3107]
		putative
		/codon_start=1
		/protein_id="BAC28852.1"
		/db_xref="GI:26330244"
		/translation="WATERVNGNGTEPEMDTTSVAIHSENFOTLLDAGLPQKVAEKLD EIVAGLVASHDLDERAIEALKEFNEDEGALVLOQFDKSDLSHVQNPDSVYSGQPSV ROREKGTQVADSSKGPDEAKIKALLERTGYTLDTVTGQKYGPPDSVYSGQPSV GTEIFVKIPLDFEDELPLFEKAGPIWDLRLMDPLTGLARGVAFVFTCTKEAAGE AVKLYNHHSIRSGKHGVCISVANNKLFVGSIPKSKTEQILIEFSKYTEGLTDVLY AKVKLVFNRLANTVTEELKESFSQFGLKRYKKLKYAFIHFDRDGAIVKAMEBN KDLEGENTIEIVAKPPDKRKAQROAKNQYDDYYFYYPHMPPTTRGRGG RGSGYPPDYGYEDYDYDYHNYRGYEDPYGYEDFVGARGRGRGARGAAP SRGGAAPRGRAGYSQSGPSRGVARGAQAQOGRGQGGKGVGEAGPDLQ"
BASE COUNT	1164 a 662 c 843 g 1169 t	
ORIGIN		
Query Match	89.0%	Score 1850.4; DB 11; Length 3838;
Best Local Similarity	94.4%	Pred. No. 0;
Matches 1966; Conservative	0; Mismatches	76; Indels 40; Gaps 3;
Qy	38	GTGGAACATCGATCGGTTTCTGCCAGCGCGTGTGAGCTTCGGCGCGCATTTTCAACAGC 97
Db	2	GTGGAACATCGATCGGTTTCTGCCAGCGCGTGTGAGCTTCGGCGCGCATTTTCAACAGC 61
Qy	98	TCACATCGCGCGGACACAGGAGGAGCGGTTTCCCGCAACCGCATACCATCG 157
Db	62	TCCATCTCGCGCGGACACAGGAGGAGCGGTTTCCCGCAACCGCATACCATCG 121
Qy	158	GACAGATTCTCGGCTCAGCCCAACGGGAGATCTCTGGAACATGCTTACAGAACAT 217
Db	122	GACAGATTCTCGGCTCAGCCCAACGGGAGATCTCTGGAACATGCTTACAGAACAT 181
Qy	218	GTTAATGGAATCGTACTGAAGAGCCCATGGATACCTTCTCGAGTTATCCATTGAA 277
Db	182	GTTAATGGAATCGTACTGAAGAGCCCATGGATACCTTCTCGAGTTATCCATTGAA 241
Qy	278	AATTTTCAGACATGCTTGATGCTGTTTACCAACAGAAAGTTCTGAAAACCTAGATGA 337
Db	242	AATTTTCAGACATGCTTGATGCTGTTTACCAACAGAAAGTTCTGAAAACCTAGATGA 301
Qy	338	ATTACGTTTCAGCGGCTAGTTGCACATAGTGTATTTAGTGAAGAGCTATTGAAGCTTTA 397
Db	302	ATTACGTTTCAGCGGCTAGTTGCACATAGTGTATTTAGTGAAGAGCTATTGAAGCTTTA 361
Qy	398	AAAGAAATTCATGAAGAAGCGGTGATGCTGTTTCAACAGTTTAAAGACAGTGTATTC 457
Db	362	AAAGAAATTCATGAAGAAGCGGTGATGCTGTTTCAACAGTTTAAAGACAGTGTATTC 421
Qy	458	TCTCATGTTTCAGAACAAAAGTGCCTTTTATGTGGAGTCAATGAGCTTTACAGCGAGA 517
Db	422	TCTCATGTTTCAGAACAAAAGTGCCTTTTATGTGGAGTCAATGAGCTTTACAGCGAGA 481
Qy	518	GAAGAAACAGGAGCAAGTAGCAGATTTCTAGTAAAGAGCAGATGAGGCAAAAATTAAG 577
Db	482	GAAGAAACAGGAGCAAGTAGCAGATTTCTAGTAAAGAGCAGATGAGGCAAAAATTAAG 541
Qy	578	GCATCTTGGAAAGAACAGGCTACACATTTGATGTGACCACTGGACAGAGAGATATGGA 637

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 00:06:34 ; Search time 2986 Seconds  
(without alignments)  
16921.971 Million cell updates/sec

Title: US-09-978-242-2

Perfect score: 2079

Sequence: 1 GGGCGCGCGCGCACCGGG.....GTTATTAACTAAAGCTACT 2079

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_estc:\*

9: gb\_estl:\*

10: gb\_est2:\*

11: gb\_est3:\*

12: gb\_est4:\*

13: gb\_est5:\*

14: gb\_est6:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1850.4	89.0	3838	11 AK034845	AK034845 Mus muscu
2	1792.8	86.2	2141	11 AK076026	AK076026 Mus muscu
3	1787.4	86.0	2030	11 AK077588	AK077588 Mus muscu
4	1537	73.9	4064	11 AK083398	AK083398 Mus muscu

5	1494.8	71.9	3289	11 AK078158	AK078158 Mus muscu
6	1386	66.7	1405	11 BC021932	BC021932 Homo sapi
7	1353.6	65.1	1361	11 BC024283	BC024283 Homo sapi
8	1343.6	64.6	2618	11 BC019360	BC019360 Homo sapi
9	1331	64.0	1359	11 BC009176	BC009176 Homo sapi
10	1327.4	63.8	1343	11 BC015575	BC015575 Homo sapi
11	983.8	47.3	1201	9 AL541211	AL541211 AL541211
12	975.2	46.9	1201	9 AL541923	AL541923 AL541923
13	921.6	44.3	1201	9 AL514248	AL514248 AL514248
14	884.2	42.5	2663	11 BC026850	BC026850 Mus muscu
15	883	42.5	1144	12 BM466029	BM466029 AGENCOURT
16	866.6	41.7	946	13 BQ686098	BQ686098 AGENCOURT
17	862.8	41.5	1080	12 BM809311	BM809311 AGENCOURT
18	844.6	40.6	905	13 BQ945048	BQ945048 AGENCOURT
19	838.2	40.3	1201	13 BX420656	BX420656 BX420656
20	836.6	40.2	909	13 BU506813	BU506813 AGENCOURT
21	831.8	40.0	868	13 BU194176	BU194176 AGENCOURT
22	829.2	39.9	910	13 BU183896	BU183896 AGENCOURT
23	826.8	39.8	879	12 BM451217	BM451217 AGENCOURT
24	807.6	38.8	929	13 BU845727	BU845727 AGENCOURT
25	768	36.9	1063	12 BM543159	BM543159 AGENCOURT
26	757.8	36.5	856	13 BQ680184	BQ680184 AGENCOURT
27	756.6	36.4	921	13 BU505018	BU505018 AGENCOURT
28	747.6	36.0	942	10 BE731663	BE731663 601567050
29	744.6	35.8	878	13 BQ650432	BQ650432 AGENCOURT
30	740.8	35.6	903	13 BQ680994	BQ680994 AGENCOURT
31	740.2	35.6	933	13 BQ950249	BQ950249 AGENCOURT
32	736.2	35.4	920	13 BQ959242	BQ959242 AGENCOURT
33	729	35.1	988	9 AV298691	AV298691 AV298691
34	728.6	35.0	891	14 CB182506	CB182506 AGENCOURT
35	721.2	34.7	826	10 BE795387	BE795387 601592924
36	719	34.6	777	13 BQ443866	BQ443866 UI-M-EW0-
37	715.6	34.4	1032	10 BE407142	BE407142 601301856
38	712.8	34.3	950	10 BG026689	BG026689 602233533
39	710.6	34.2	798	12 BI558562	BI558562 603240561
40	709.8	34.1	823	10 BG282140	BG282140 602403281
41	704.6	33.9	772	13 BQ571781	BQ571781 UI-M-FC0-
42	700.4	33.7	1088	12 BI661316	BI661316 603304285
43	700.2	33.7	860	10 BE560849	BE560849 601346126
44	698.8	33.6	895	12 BI558142	BI558142 603240061
45	697.8	33.6	856	13 BU191822	BU191822 AGENCOURT

## ALIGNMENTS

RESULT 1	AK034845	AK034845	3838 bp	linear	HTC 05-DEC-2002
LOCUS	Mus musculus	Mus musculus	12 days embryo	embryonic body	between diaphragm region
DEFINITION	and neck CDNA, RIKEN full-length enriched library, clone:9430046J23	product: NSI-associated protein 1, full insert sequence.			
ACCESSION	AK034845	AK034845	1	GI:263330243	
VERSION	AK034845	AK034845	1	GI:263330243	
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
MEDLINE	99279253				
PUBMED	10349636				
REFERENCE	2				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	20499374				
PUBMED	11042159				